Fall 11-27-2017

POLYPLOIDY AND HYBRIDIZATION HAVE A ROLE IN PLANT DISTRIBUTION IN THE UNITED STATES: A BIOINFORMATICS STUDY

Andrea Ortiz
University of Texas at Tyler

POLYPLOIDY AND HYBRIDIZATION HAVE A ROLE IN PLANT DISTRIBUTION IN THE UNITED STATES: A BIOINFORMATICS STUDY

by

ANDREA ORTIZ

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Biology

Department of Biology

Kate Hertweck, Ph.D., Committee Chair

College of Science

The University of Texas at Tyler
December 2017

The University of Texas at Tyler

Tyler, Texas

This is to certify that the Master's Thesis Dissertation of

ANDREA ORTIZ

has been approved for the thesis/dissertation requirement on

October 19

for the Biology degree

Approvals:

Thesis Chair: Kate Hertweck, Ph.D.

Member: Josh Banta, Ph.D.

Member: Katrin Kellner, Ph.D.

Chair: Lance Williams, Department of Biology

Dean, College of Arts and Sciences

Table of Contents

List of Tables……..…....iv

List of Figures..…….....…….....v

Abstract…….……viii

Chapter 1 Introduction…….....1

Niche modeling……..……....…1

Model Systems……...2

Callisia graminea…….……...……...2

Allium canadense…….……...3

Chapter 2 Methods…….…..5

Locality and weather data……..5

Building Niche Models……..6

Comparisons of Niche Model Overlap…….……..7

Chapter 3 Results…...…….…….……9

Aim 1…….…….9

Aim 2…….…….9

Aim 3…….…….9

Aim 4…….…….....11

Aim 5…….……...11

Comparisons of Niche Model Overlap…….....14

Chapter 4 Discussion..……...40

Chapter 5 Bibliography…...……...……...43

List of Tables

Table 1 *C. graminea* niche and range comparisons……..…...7

Table 2 *A. canadense* range comparisons…….……..….8

Table 3 Results of *C. graminea* niche comparisons…….….14

Table 4 Test gain for *C. graminea* cytotypes for 1929…….....…16

Table 5 Test gain or *C. graminea* cytotypes for 2011……...…16

Table 6 Results of *A. canadense* range comparisons…….27

Table 7 Test gain for *A. canadense* varieties for 1929……...…...28

Table 8 Test gain for *A. canadense* varieties for 2011……..28

List of Figures

Figure 1. Distribution of *C. graminea* cytotypes……..……..2

Figure 2. Distribution of *A. canadense* varieties……....3

Figure 3. Models of *C. graminea* diploids in 1929 and *C. graminea* diploids in 2011…….…..……..15

Figure 4. Models of *C. graminea* tetraploids in 1929 and *C. graminea* tetraploids in 2011……....15

Figure 5. Combined Models of *C. graminea* diploids and tetraploids in 1929 and *C. graminea* diploids and tetraploids in 2011.....16

Figure 6. Response curves for *C. graminea* diploids for 1929…...…....17

Figure 7. Response curves for *C. graminea* tetraploids for 1929……...17

Figure 8. Response curves for *C. graminea* both cytotypes for 1929…..18

Figure 9. Response curves for *C. graminea* diploids for 2011……....19

Figure 10. Response curves for *C. graminea* tetraploids for 2011…..…..20

Figure 11. Response curves for *C. graminea* both cytotypes for 2011…..21

Figure 12. A plot of the Principal Components Analysis (PCA) for *C. graminea* cytotypes for 1929.....…….22

Figure 13. A plot of the Principal Components Analysis (PCA) for *C. graminea* cytotypes for 2011.....….…….23

Figure 14. Models of *A. c. canadense* in 1929 and *A. c. canadense* in 2011…….....23

Figure 15. Models of *A. c. lavendulare* in 1929 and *A. c. lavendulare* in 2011..……...…….....24

Figure 16. Models of *A. c. mobilense* in 1929 and *A. c. mobilense* in 2011.....……...…..24

Figure 17. Models of *A. c. fraseri* in 1929 and *A. c. fraseri* in 2011.....…….....25

Figure 18. Models of *A. canadense* parentals in 1929 and *A. canadense* hybrids in 1929 .....……...……..25

Figure 19. Models of *A. canadense* parentals in 2011 and *A. canadense* hybrids in 2011.....……..……26

Figure 20. Models of *A. canadense* parentals in 1929 and *A. canadense* parentals in 2011 ..... ……26

Figure 21. Models of *A. canadense* hybrids in 1929 and *A. canadense* hybrids in 2011 .....……..27

Figure 22. Response curves for *A. c. canadense* for 1929……...…….....28

Figure 23. Response curves for *A. c. lavendulare* for 1929…...…...……29

Figure 24. Response curves for *A. c. mobilense* for 1929…….……29

Figure 25. Response curves for *A. c. fraseri* for 1929……..……30

Figure 26. Response curves for *A. canadense* parentals for 1929……..…30

Figure 27. Response curves for *A. canadense* hybrids for 1929……...…31

Figure 28. Response curves for *A. c. canadense* for 2011…...……..…32

Figure 29. Response curves for *A. c. lavendulare* for 2011...……33

Figure 30. Response curves for *A. c. mobilense* for 2011...……..…..34

Figure 31. Response curves for *A. c. fraseri* for 2011……...……35

Figure 32. Response curves for *A. canadense* parentals for 2011…....…36

Figure 33. Response curves for *A. canadense* hybrids for 2011……...…37

Figure 34. A plot of the the Principal Components Analysis (PCA) for *A. canadense* varieties for 1929.....……..…38

Figure 35. A plot of the the Principal Components Analysis (PCA) for *A. canadense* varieties for 2011.....39

Abstract

POLYPLOIDY AND HYBRIDIZATION HAVE A ROLE IN PLANT DISTRIBUTION IN THE UNITED STATES: A BIOINFORMATICS STUDY

Andrea Ortiz

Thesis Chair: Kate Hertweck, Ph.D.

The University of Texas at Tyler
December 2017

Introduction: Polyploidy is the doubling of a genome. Autopolyploidy and allopolyploidy are two different modes of genome duplication. These events are common in angiosperms and usually involve diploids and tetraploids, a resulting polyploid species. Both cytotypes exhibit competitive exclusion and environmental adaptation according to their spatial distributions. My research focuses on how abiotic environmental factors affect the range and the amount of range overlap of the taxon within the *Callisia graminea* and *Allium canadense* systems over time. Niche modeling determines whether there is substantial niche overlap between the *Callisia graminea* diploid and tetraploid, whether range shifts occur between the *Callisia graminea* cytotypes or *Allium canadense* varieties, and whether there is substantial range overlap between the *Allium canadense* parental and hybrid varieties.

Methods: The environmental layers from 1929 and 2011 were used for building the niche models and were downloaded from PRISM (http://www.prism.oregonstate.edu/recent/). The layers were clipped to encompass the range that each system occupies. The niche models were constructed using MaxEnt and assessed for robustness. Area under the receiver operating characteristic curve values were utilized to evaluate the accuracy of each model. The model-based methods used to evaluate niche overlap were Schoener's D and the I Similarity tests and permutation tests. The other statistic used to evaluate niche overlap was Principal Component Analysis (PCA).

Key Results: For the *Callisia graminea* system, the range overlap of the diploids and tetraploids increased when comparing the 1929 models to the 2011 models. Surprisingly, diploid range increased, while the tetraploid range decreased. For the *Allium canadense* system, the range of variety *lavendulare* increased in 2011. Though the range of hybrid varieties had decreased in 2011, it was larger compared to the range of the parental

varieties of the same year. The parentals' range increased when comparing the 1929 model to the 2011 model.

Conclusions: The increase in niche overlap over time between the *Callisia graminea* cytotypes indicates that the two cytotypes are not diverging ecologically, but actually appear to be converging ecologically instead. The *Callisia graminea* tetraploids and the *Allium canadense* hybrids have experienced its boundary of accommodating environmental conditions and may not be capable of expanding further, whereas the *Callisia graminea* diploids and the *Allium canadense* parentals have not. The variety *lavendulare* has successfully expanded by encountering favorable abiotic conditions.

Chapter One
Introduction

Polyploidy, or genome duplication, has enabled genetic innovation that fueled the evolution of angiosperms (Soltis et al., 2009; Jiao et al., 2011; Magadum et al., 2013; Mühlhausen and Kollmar, 2013; Glennon et al., 2014; Soltis et al., 2014). Gene duplication occurs by autopolyploidy or allopolyploidy (Ramsey and Ramsey, 2014). Whole genome duplication events may give rise to instant speciation which increases biodiversity (Soltis et al., 2014). Diploid progenitors and resulting tetraploids are common cytotypes that display reproductive isolation due to the tendency of tetraploids to demonstrate self-fertilization (Ramsey and Ramsey, 2014). Both cytotypes exhibit competitive exclusion and environmental adaptation according to their spatial distributions. Niche modeling can be used to characterize spatial distributions for groups within species. My research focuses on how abiotic environmental factors affect the range and the amount of range overlap of taxa within the *Callisia graminea* and *Allium canadense* systems over time. Niche modeling determines whether there is substantial niche overlap between the *Callisia graminea* diploid and tetraploid, whether range shifts occur between the *Callisia graminea* cytotypes or *Allium canadense* varieties, and whether there is substantial range overlap between *Allium canadense* parental and hybrid varieties. The statistics used to evaluate niche overlap were Principal Component Analysis, Schoener's D and the I Similarity tests, and permutation tests. The increase in niche overlap over time between the *Callisia graminea* cytotypes indicates that the two cytotypes are not diverging ecologically, but actually appear to be converging ecologically instead. The *Callisia graminea* tetraploids and the *Allium canadense* hybrids have experienced its boundary of accommodating environmental conditions and may not be capable of expanding further, whereas the *Callisia graminea* diploids and the *Allium canadense* parentals have not. Variety *lavendulare* has successfully expanded by encountering favorable abiotic conditions.

Niche modeling

A niche is defined as a characteristic of a population or species regarding abiotic and biotic environment factors (Colwell and Rangel, 2009). The factors influencing the establishment, retraction, or expansion of a taxon's range can be seen when studying its niche and how it changed over time (Warren et al., 2008). I will only be assessing the environmental space which contains the known species distribution and the abiotic environmental factors that contribute to its success (Peterson, 2006; Mcinerny and Etienne, 2012; Glennon et al., 2014). For plants those factors may include (but are not limited to) precipitation, temperature, and soil type. Ecological niche modeling can assist conservation managers on best management practices for threatened species, assess the effects of climate change on a population, or evaluate the potential spread of an invasive species (Allouche et al., 2006). When collecting in the field, data collected may be incomplete, biased, incorrectly identified, missing biotic and mobility factors, and/or indicate misleading dispersal patterns (Warren, 2012). Ecological niche modeling is robust to these problems (Warren, 2012).

Model Systems

I used ecological niche modeling to study the niche evolution of *C. graminea* cytotypes and the range evolution of *A. canadense* varieties to investigate the relationship between these organisms and their environment over time. Both systems are composed of species that are monocotyledons and have had no human interference (Giles, 1942). These plant systems have not been subjected to cultivation which can make them difficult to work with when artificial changes have been made to the organisms (Giles, 1942). These two plant complexes exhibit polyploidy but the *A. canadense* system also utilizes hybridization. Sexual reproduction is seen in all varieties of *A. canadense* except the asexual var. *canadense* (Wheeler, 2011). In the *C. graminea* system, the tetraploid has a larger geographic range than the diploid (Giles, 1942; Giles 1943).

Callisia graminea

Callisia graminea (Small) G. Tucker belongs to the plant family Commelinaceae characterized as a herbaceous, perennial with glabrous roots and pink to rose colored petals (Giles 1942). It occurs in both diploid and autotetraploid populations in the Southeast United States (Figure 1; Giles 1942; Giles 1943). This autotetraploid seems to be characterized by larger, more vigorous plants, greater ecological amplitude, and a distinct, larger geographic range compared to its diploid progenitor, indicating there may be an advantage to possessing a duplicated genome (Giles 1942; Giles 1943).

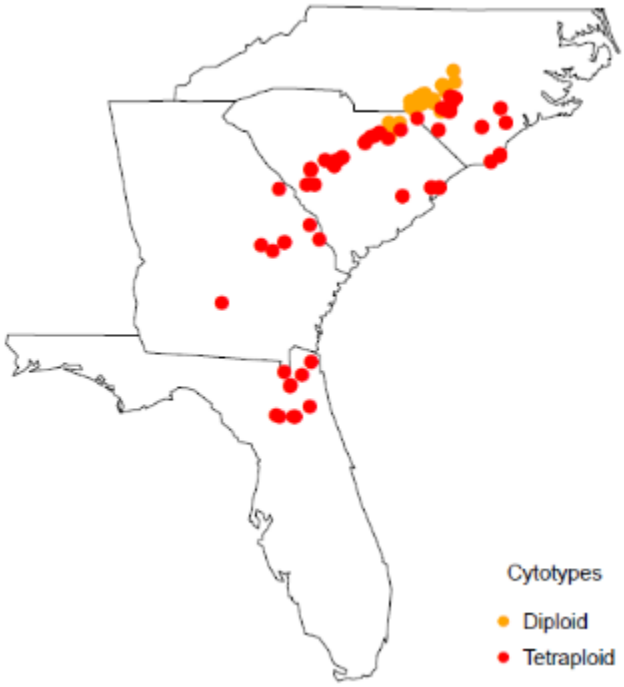


Figure 1. Distribution of *C. graminea* cytotypes.

Allium canadense

Allium canadense L. is in the family Amaryllidaceae and is characterized as herbaceous, xerophytic, and possessing true bulbs (Wheeler, 2011). Polyploidy and hybridization have resulted in six varieties comprising the *A. canadense* complex (Ownbey, 1955; Wheeler, 2011). It is suggested that the diploid varieties *A. c. fraseri* and *A. c. mobilense* hybridized in three different geographic locations, via allopolyploidy, to give rise to diploid varieties *ecristatum*, *hyacinthoides*, and *lavendulare* (Figure 2; Wheeler, 2011; Ramsey and Ramsey, 2014). Most of the varieties are sexually reproducing, with the exception of the var. *A. c. canadense* which is asexual (Ownbey 1955; Wheeler, 2011). *A. c. canadense* is a tetraploid that has formed three separate times from *A. c. fraseri*, *mobilense*, and *lavendulare* (Wheeler, 2011). The only other polyploid is seen in var. *A. c. lavendulare* (Wheeler, 2011).

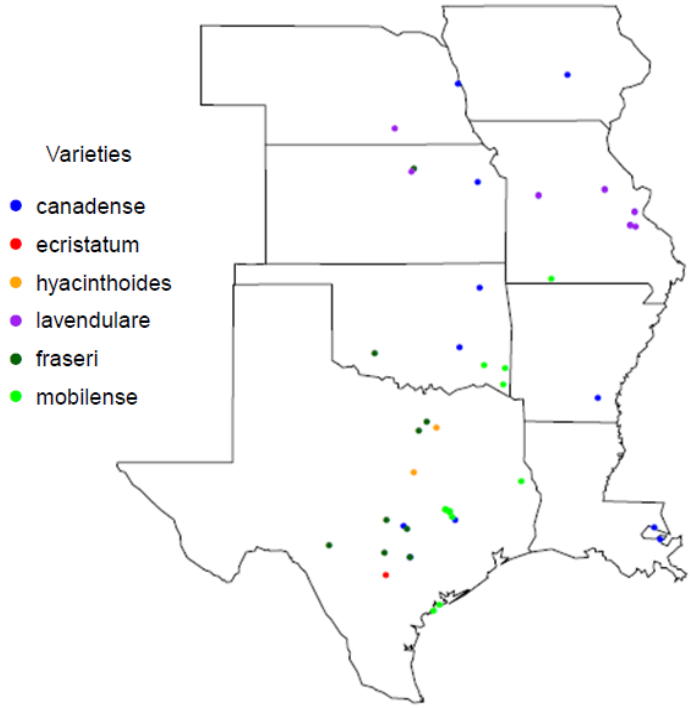


Figure 2. Distribution of *A. canadense* varieties. The var. *A. c. mobilense* and var. *A. c. fraseri* are the parentals varieties whereas the var. *A. c. ecristatum*, var. *A. c. lavendulare* and var. *A. c. hyacinthoidies* are hybrids varieties. Only var. *A. c. canadense* reproduces asexually.

The goal of my thesis is to determine how the *C. graminea* system's cytotypes then the *A. canadense* system's varieties spatially separated and interacted with one another over time and observe the effect on niche overlap within each plant system. Ecological niche modeling will assist in establishing which climatic factors affect plant distribution by using historical occurrence data and visualizing the induced change in the range of these two plant systems when comparing the historical and contemporary climate models.

Aims:

- Aim 1: Enable the replication of my range analyses by publishing scripts on GitHub. I will publish my scripts on a freely available, public repository which will enable my material to be used by others in their range analyses.

- Aim 2: Assess historical and contemporary niche models for diploid and tetraploid *C. graminea* and test for overlap between cytotypes. Their additional genetic material can cause the ecological amplitudes of tetraploids to differ from their diploid progenitors that enable them access to areas that would otherwise be off limits (Levin, 2003). I expect substantial niche overlap in the historical *C. graminea* cytotype models, with a decrease of overlap in contemporary models.

- Aim 3 & 4: Identify whether range shifts for *C. graminea* cytotypes (Aim 3) or *A. canadense* varieties (Aim 4) may have occurred between the historical and contemporary models. Temperature and precipitation have a major influence on plant physiology, setting the limitations and adaptability of a species and establishing its geographic boundaries (Theodoridis et al., 2013). When a species is first introduced to an environment it has a narrow range but with time that range expands (Johnson and Ashman, 2014; Thompson et al., 2014). I expect the *C. graminea* tetraploids range and the variety *A. c. lavendulare*, which has a diploid and tetraploid population, range to expand over time. I expect that the *C. graminea* diploids range will remain constant and the var. *A. c. lavendulare* range will not remain consistent.

- Aim 5: Assess historical and contemporary niche models for *A. canadense* parental and hybrid varieties. Hybridization increases biodiversity which enables the hybrids to be more robust and colonize areas that the parentals cannot (Levin, 2003). I expect that the parental varieties *A. c. mobilense* and *A. c. fraseri* would have ranges that remain consistent whereas the hybrid varieties *A. c. lavendulare*, *A. c. ecristatum*, and *A. c. hyacinthoides* will expand.

Chapter Two
Methods

Locality and weather data-The locality data for *C. graminea* cytotypes and *A. canadense* varieties were obtained from Giles (1942; 1943) and Ownbey (1955), respectively. I used Google Maps to assign the locality description of each specimen a latitude/longitude format. I used R (R Core Team, 2014) in RStudio (RStudio Team, 2015) to perform all subsequent analysis; code is freely available at https://github.com/aortiz24/Callisia_nicheModeling and https://github.com/aortiz24/Allium_nicheModeling. I used R packages dplyr (Wickham and Francois, 2015) and biomod2 (Thuiller et al., 2014) for data parsing. I downloaded a shapefile from the United States Census Bureau (http://www2.census.gov/geo/tiger/GENZ2015/shp/) to obtain the outlines of US states. All occurrence points were plotted onto the shapefile for taxon ranges and any outliers that were offset in respect to the shapefile's boundaries were removed.

I downloaded and clipped seven historical and contemporary climate layers from the PRISM Climate Group (http://www.prism.oregonstate.edu/recent/) to the extent of the shapefile (Hijmans, 2015). These spatial climate datasets are generated by gathering climate observations from monitoring networks and are subjected to quality control. Then they can be used in various modeling techniques and cover different time periods such as historical past (1895 - 1980) and recent years (January 1981 - February 2017) where I obtained the climate layers used in this study. I determined that the PRISM 1929 layers, used for the historical models, and the PRISM 2011 layers, used for the contemporary models, were not anomalous using the time series graphs constructed at the National Centers for Environmental Information (https://www.ncdc.noaa.gov/cag/). Any conclusions drawn from the models would not be biased due to extreme weather conditions.

I performed two separate Pearson correlation analyses each for *C. graminea* and *A. canadense*. Highly correlated layers can interfere with the interpretability of the species distribution model (Yong et. al., 2013; Merow et. al., 2013). When there were climate layers correlated at more than 0.7, the set of layers used for modeling was reduced to eliminate redundant variables causing these correlations. The historical model for *C. graminea* was built with three layers from the historical past which were: mean temperature, precipitation, and minimum vapor deficit. The contemporary model for *C. graminea* used four layers from recent years, which were: mean temperature, precipitation, minimum vapor pressure deficit, and maximum vapor pressure deficit. The historical model for *A. canadense* was built with three layers from the historical past, which were: mean temperature, precipitation, and maximum vapor deficit. The contemporary model for *A. canadense* used four layers from recent years, which were: mean temperature, precipitation, minimum vapor pressure deficit, and mean dewpoint temperature.

Building Niche Models-I generated PRISM 1929 and PRISM 2011 niche models using MaxEnt (Hijmans et al., 2015) for each taxon and R packages raster (Hijmans, 2015) and dismo (Hijmans et al., 2015). For the *Callisia graminea* PRISM 1929 & 2011 models, diploids had 31 occurrence points and tetraploids had 79 occurrence points, for a total of 114 occurrence points. For the *Allium canadense* PRISM 1929 & 2011 models, *canadense* had 10 occurrence points, *mobilense* had 13 occurrence points, *fraseri* had 9 occurrence points, *lavendulare* had 6 occurrence points, *hyacinthoides* had 2 occurrence points, and *ecristatum* had 1 occurrence point, 41 total occurrence points for parentals and 41 for hybrids.

MaxEnt takes the list of species occurrence points, known a presence-only data, and extracts the information from the environmental layers at those specific points (Merow et al., 2013). Then, the program selects random locations on the environmental layers and extracts information for these background points (Merow et al., 2013). The program compares information for both of these points and is able to see a preference in conditions of the species and can construct a model (Merow et al., 2013). Each taxon model had specific parameters such as testing and training sets for each taxon, cross-validation of models, random seed, background points, test gain, and response curves (Hijmans et al., 2015). The two data sets were assigned by randomly splitting 80% of the species occurrence data to the training data set and the remaining 20% to the testing data set for taxa which had greater than 30 occurrence points. For taxa that had less than 30 occurrence points, the leave-one-out method was used. For this method, the number of folds was equal to the number of occurrence points for each taxon so that each fold contained one less than the total occurrence points in order to include one test data point in each fold. The folds were then averaged together for each taxon (Walters et al., 2017). Random seed nonspecifically selects a new set of species occurrence points every time MaxEnt is run (Phillips, 2010; Hijmans et al., 2015). Background points are randomly taken to represent the environmental conditions of the study area (Phillips et al., 2006; Hijmans and Elith, 2017). Test gain is used to measure how well a model fits the data it is given by comparing the model's performance to another model that allocated the same habitat suitability score through the landscape (Walters et al., 2017; Phillips, 2006). The climate variable with the highest test gain that is closest to the test gain of the full model is the variable that is most important to model predictability (Walters et al., 2017). Response curves show the relationship between the climatic variable and the suitability of the model when all other climatic variable are held constant (Phillips, 2010). I used the area under the receiver operating characteristic curve (AUC value) to assess the accuracy of the PRISM 1929 and PRISM 2011 predictive models (Thompson et al., 2014). The Area under the Operator Receiving curve (AUC) measures the probability that the habitat suitability of a randomly chosen presence site will be higher than a randomly chosen pseudoabsence point (Walters et al., 2017; Phillips & Dudik, 2008). The test AUC refers to the average ratio of the pseudoabsence data with lower habitat suitability scores for a single 'test' presence locations left out of the model building process for each model replication (Walters et al., 2017; Phillips & Dudik, 2008).All of the models had test AUC values greater than 0.75 (Table 4) and is sufficient for the prediction of species occurrence (Walters et al., 2017; Elith, 2002). I generated predictive distribution maps of

C. graminea cytotypes and *A. canadense* varieties using taxa occurrence data, and both historical and contemporary layers using MaxEnt in R (Tables 1 and 2).

Comparisons of Niche Model Overlap-I assessed the overlap of ranges from various cytotype and from variety comparisons using both standard statistics and model-based methods (Table 1; Table 2). The statistical methods I utilized are the Principal Component Analysis (PCA), which determined the environmental variables that are responsible for simulating each taxon's suitable habitat (Glennon et al., 2014). ENMTools was implemented via the ENMeval package in R using two model-based metrics, Schoener's D and the I Similarity tests (Warren et al., 2008; Muscarella et al., 2014; Hijmans et al., 2015). I performed permutation tests in both ENMTools (Phillips et al., 2006; Warren et al., 2008; Nakazato et al., 2010; Warren et al., 2010; Glor et al., 2011; Warren and Seifert, 2011) and the ENMTools in ENMeval R package within R statistical programming (Muscarella et al., 2014) which used the same criteria of each model built, except it altered the labels of the occurrence points. The test takes the occurrence data and randomly misidentifies the occurrence point coordinates with incorrect labels (Phillips et al., 2006; Warren et al., 2008; Nakazato et al., 2010; Warren et al., 2010; Geange et al., 2011; Glor et al., 2011; Warren and Seifert, 2011). Then, MaxEnt models are constructed with the misidentified occurrence points.

This process is repeated 100 times along with the niche overlap statistic of each model (Phillips et al., 2006; Warren et al., 2008; Nakazato et al., 2010; Warren et al., 2010; Geange et al., 2011; Glor et al., 2011; Warren and Seifert, 2011). The niche overlap statistic that is fifth lowest represents the threshold of significance which is the same as the p value equals 0.05. The claim that two ranges are significantly different from one another cannot be reached unless the niche overlap values produced by the original data sets are lower than the critical value produced by the permuted data sets.

Table 1: *C. graminea* range comparisons. "Combined" means both diploid (2x) and tetraploid (4x) are included in the same model.

Related Aim	Layers	Cytotypes
Aim 2	PRISM1929	2x vs 4x
Aim 2	PRISM2011	2x vs 4x
Aim 3	PRISM1929 vs PRISM2011	2x
Aim 3	PRISM1929 vs PRISM2011	4x
Aim 3	PRISM1929 vs PRISM2011	combined

Table 2: *A. canadense* range comparisons. "Combined" means all six varieties are included in the same model.

Related Aim	Layers	Varieties
Aim 4	PRISM1929 vs PRISM2011	*lavendulare*
Aim 4	PRISM1929 vs PRISM2011	*canadense*
Aim 4	PRISM1929 vs PRISM2011	*ecristatum*
Aim 4	PRISM1929 vs PRISM2011	*mobilense*
Aim 4	PRISM1929 vs PRISM2011	*fraseri*
Aim 5	PRISM1929 vs PRISM2011	parental (*mobilense*, *fraseri*) vs hybrid (*ecristatum*, *lavendulare*, *hyacinthoides*)

Chapter Three
Results

Aim 1: I have published my scripts on a freely available, public repository on GitHub (https://github.com/aortiz24/Callisia_nicheModeling; https://github.com/aortiz24/Allium_nicheModeling) which will enable my material to be used by others in their analyses.

***Aim 2 : Callisia* diploid versus tetraploid comparisons**

I expected there would be substantial niche overlap in the historical *C. graminea* cytotype models, with a decrease of overlap in contemporary models. The comparison between diploids and tetraploids in 1929 determined that their ranges were significantly different from each other (Table 3). The comparison between diploids and tetraploids in 2011 determined that their ranges were significantly different from each other (Table 3). The substantial niche overlap for the historical comparison of *C. graminea* cytotype models was 0.78 (Table 3). Contrary to expectation, there was an increase in niche overlap in the contemporary models (Table 3).

***Aim 3: Callisia* diploid and tetraploid comparisons**

I expected the *C. graminea* tetraploid range would expand over time and the *C. graminea* diploid range would remain constant. Diploid ranges in 1929 and 2011 were significantly different from each other (Table 3). Diploid range did not remain constant when comparing 1929 and 2011 (Figure 3). Tetraploid ranges in 1929 and 2011 were not significantly different from each other (Table 3). Tetraploid range decreased when comparing 1929 and 2011 (Figure 4). Combined ranges in 1929 and 2011 were not significantly different from each other (Table 3). The combined model for 1929 and 2011 expanded (Figure 5). Ecological tolerances such as temperature, salinity, drought, herbivory and shade vary among cytotypes and can support or restrict the growth of the *C. graminea* tetraploid establishing geographic boundaries, in this case, creating a patchy distribution (Levin, 2003; Theodoridis et al., 2013; Figure 4).

All of the *C. graminea* models from 1929 and 2011 had test AUC values greater than 0.75 (Table 4; Table 5) and are sufficient for the prediction of species occurrence (Walters et al., 2017; Elith, 2002). Mean temperature was most important for the *C. graminea* cytotypes in 1929 (Table 4). Minimum vapor pressure deficit was most important for diploids in 2011 (Table 5). Mean temperature was most important for tetraploids and the model including occurrence data of both cytotypes in 2011 (Table 5). Interestingly, the *C. graminea* diploid model responded to the minimum vapor pressure deficit, or the minimal difference in the total air moisture compared to the actual air moisture, indicating a possible difference in how this cytotype interacts with its environment compared to the tetraploid model and the model including occurrence data of both cytotypes (Table 5; Prince et al., 1998).

The response curves for the *C. graminea* diploids in 1929 indicate that the most suitable habitat had the following characteristics: (a) mean temperature between 14 to 19 degrees

Celsius (Figure 6a); (b) precipitation between 1300 to 1750 mm (Figure 6b); (c) maximum vapor pressure deficit between 0 to 4 (Figure 6c).

The response curves for the *C. graminea* teraploids in 1929 indicate that the most suitable habitat had the following characteristics: (a) mean temperature between 11 to 25 degrees Celsius (Figure 7a); (b) precipitation steadily decreasing between 1300 to 1900 mm (Figure 7b); (c) maximum vapor pressure deficit between 0 to 5 (Figure 7c).

The response curves for both cytotypes of *C. graminea* in 1929 indicate that the most suitable habitat had the following characteristics: (a) mean temperature between 14 to 25 degrees Celsius (Figure 8a); (b) precipitation between 1300 to 1900 mm (Figure 8b); (c) maximum vapor pressure deficit between 1 to 3 (Figure 8c).

The response curves for the *C. graminea* diploids in 2011 indicate that the most suitable habitat had the following characteristics: (a) mean temperature steadily decreasing between 0 to 18 degrees Celsius (Figure 9a); (b) precipitation steadily decreasing between 0 to 1300 mm (Figure 9b); (c) minimum vapor pressure deficit between 17 to 21 (Figure 9c); (d) mean dewpoint temperature between 0 to 3 (Figure 9d).

The response curves for the *C. graminea* diploids in 2011 indicate that the most suitable habitat had the following characteristics: (a) mean temperature between 14 to 26 degrees Celsius (Figure 10a); (b) precipitation steadily decreasing between 0 to 1300 mm (Figure 10b); (c) minimum vapor pressure deficit steadily decreasing between 19 to 25 (Figure 10c); (d) mean dewpoint temperature steadily decreasing between -1 to 4 (Figure 10d).

The response curves for the *C. graminea* diploids in 2011 indicate that the most suitable habitat had the following characteristics: (a) mean temperature between 15 to 27 degrees Celsius (Figure 11a); (b) precipitation steadily decreasing between 800 to 1300 mm (Figure 11b); (c) minimum vapor pressure deficit between 13 to 25 (Figure 11c); (d) mean dewpoint temperature steadily increasing between 0 to 5 (Figure 11d).

The Principal Components Analysis (PCA) showed two components contribute to 84.7% of the total variation among the two *Callisia* cytotypes (PC1= 63.6%, PC2= 21.1%; Figure 12). Precipitation and maximum vapor pressure deficit equally contributed to PC1 and PC2 (Figure 12). Mean temperature is more strongly associated with PC1 (Figure 12). The Principal Component Analysis (PCA) and test gain determined that mean temperature is an important environmental variable in the model for *C. graminea* cytotypes in 1929 (Figure 12; Table 4). The PCA also indicated that the cytotypes are not overlapping which suggests the cytotypes occupy distinct ranges in 1929 (Figure 12).

The Principal Component Analysis showed that two components contribute to 82.9% of the total variation among the two *C. graminea* cytotypes (PC1= 53.1%, PC2= 29.8%; Figure 13). Precipitation and minimum vapor pressure deficit equally contributed to PC1 and PC2 (Figure 13). Mean temperature is more strongly associated with PC1 (Figure 13). The Principal Component Analysis (PCA) and test gain determined that mean temperature is an important environmental variable used to simulate the suitable habitat

for *C. graminea* tetraploids and the *C. graminea* cytotypes in 2011 (Figure 13; Table 5). The PCA also indicated that the cytotypes are not overlapping which suggests the cytotypes occupy distinct ranges in 2011 (Figure 13).

***Aim 4: Allium* varieties**

I expected that the variety *A. c. lavendulare*, which has a diploid and tetraploid population, range would expand over time. I expected that the var. *A. c. lavendulare* range would not remain consistent. The *canadense* ranges in 1929 and 2011 were not significantly different from each other (Table 6). The range for *A. c. canadense* expanded when comparing 1929 and 2011 (Figure 14). The *lavendulare* ranges in 1929 and 2011 were not significantly different from each other (Table 6). The *lavendulare* range increased when comparing 1929 and 2011 (Figure 15). The *mobilense* ranges in 1929 and 2011 were not significantly different from each other (Table 6). The *mobilense* range expanded when comparing 1929 and 2011 (Figure 16). The *fraseri* range in 1929 and 2011 were not significantly different from each other (Table 6). The *fraseri* range decreased when comparing 1929 and 2011 (Figure 17). The nine occurrence points for *A. c. fraseri* were spread out across the center of Texas, Oklahoma,and Kansas (Figure 2; Figure 17). MaxEnt builds more informative models than other species modeling methods with as few as five occurrence points (Hernandez et al., 2006). If the distinction between background occurrence points is not clear, which is likely with very small sample sizes, the resulting models (such as Figure 17) are uninformative.

***Aim 5: Allium* parental and hybrid varieties**

I expected that the parental varieties *A. c. mobilense* and *A. c. fraseri* would have ranges that remain consistent whereas the hybrid varieties *A. c. lavendulare*, *A. c. ecristatum*, and *A. c. hyacinthoides* would expand. The ranges of the parentals in 1929 and hybrids in 1929 were not significantly different from each other (Table 6). The ranges of the parentals in 2011 and hybrids in 2011 were not significantly different from each other (Table 6). In the *A. canadense* system, the hybrids have a larger range than the parentals in 1929 and 2011 (Figure 18; Figure 19). The ranges of parentals in 1929 and 2011 were not significantly different from each other (Table 6). The ranges of hybrids in 1929 and 2011 were not significantly different from each other (Table 6; Figure 21). When comparing the parentals in 1929 to 2011, the range has expanded whereas the hybrid range has contracted (Figure 20: Figure 21).

All of the models except for the *fraseri* model had test AUC values greater than 0.75 (Table 7) and is sufficient for the prediction of species occurrence (Walters et al., 2017; Elith, 2002). Precipitation was most important for the *canadense*, *mobilense*, parentals, and hybrids models (Table 7). Maximum vapor pressure deficit was most important for the *lavendulare* model (Table 7). Mean temperature was most important for the *fraseri* model (Table 7). Contrary to PCA results, test gain indicates that precipitation is playing a bigger role in the *A. canadense* system since *canadense* and *mobilense* varieties are responding to it (Table 7). Test gain revealed that maximum vapor pressure deficit, or the maximum difference in the total air moisture compared to the actual air moisture, was most important only to the *lavendulare* variety (Table 7; Prince et al., 1998).

All of the models except for the *canadense*, *fraseri*, and hybrids model had test AUC values greater than 0.75 (Table 8) and is sufficient for the prediction of species occurrence (Walters et al., 2017; Elith, 2002). Minimum vapor pressure deficit was most important for the *canadense* and *lavendulare* models (Table 8). The *canadense* and *lavendulare* varieties respond to minimum vapor pressure deficit (Table 8). Mean dewpoint temperature was most important for the *mobilense* model (Table 8). The *mobilense* variety is responding to mean dewpoint temperature, which is a precise way to measure the moisture in the air to pinpoint the temperature at which dew forms (Table 8; Hubbard et al., 2003). No climatic variables were most important for the *fraseri* model (Table 8). The *fraseri* variety did not respond to any of the climatic variables included in the model (Table 8). In hindsight, *fraseri* needed to be modeled with different layers due to low predictability (AUC < 0.75) in 1929 and 2011 (Table 7; Table 8). Mean temperature was most important for the parentals model (Table 8). Precipitation was most important for the hybrids model (Table 8). For the parentals and hybrids in 2011, precipitation and mean temperature seem to be more important to the *A. canadense* system compared to the other environmental layers included in the 2011 model (Table 8).

The response curves for *A. c. canadense* in 1929 indicate that the most suitable habitat had the following characteristics: (a) mean temperature did not contribute to the predictability of the model (Figure 22a); (b) precipitation steadily increasing between 1000 to 1500 mm (Figure 22b); (c) maximum vapor pressure deficit steadily decreasing between 13 to 17 (Figure 22c).

The response curves for *A. c. lavendulare* in 1929 indicate that the most suitable habitat had the following characteristics: (a) mean temperature steadily decreasing between 10 to 14 degrees Celsius (Figure 23a); (b) precipitation steadily increasing between 1200 to 1500 mm (Figure 23b); (c) maximum vapor pressure deficit steadily decreasing between 11 to 15 (Figure 23c).

The response curves for *A. c. mobilense* in 1929 indicate that the most suitable habitat had the following characteristics: (a) mean temperature steadily increasing between 10 to 24 degrees Celsius (Figure 24a); (b) precipitation steadily increasing between 1200 to 1500 mm (Figure 24b); (c) maximum vapor pressure deficit steadily decreasing between 13 to 20 (Figure 24c).

The response curves for *A. c. fraseri* in 1929 indicate that the most suitable habitat had the following characteristics: (a) mean temperature steadily increases between 10 to 24 degrees Celsius (Figure 25a); (b) precipitation did not contribute to the predictability of the model (Figure 25b); (c) maximum vapor pressure deficit slightly increases between 11 to 40 (Figure 25c).

The response curves for *A. canadense* parentals in 1929 indicate that the most suitable habitat had the following characteristics: (a) mean temperature between 15 to 24 degrees Celsius (Figure 26a); (b) precipitation between 700 to 1300 mm (Figure 26b); (c)

maximum vapor pressure deficit did not contribute to the predictability of the model (Figure 26c).

The response curves for *A. canadense* hybrids in 1929 indicate that the most suitable habitat had the following characteristics: (a) mean temperature steadily increasing between 16 to 25 degrees Celsius (Figure 27a); (b) precipitation between 750 to 1300 mm (Figure 27b); (c) maximum vapor pressure steadily decreasing between 10 to 15 (Figure 27c).

The response curves for the *A. c. canadense* in 2011 indicate that the most suitable habitat had the following characteristics: (a) mean temperature slightly decreasing (Figure 28a); (b) precipitation steadily increasing between 250 to 1600 mm (Figure 28b); (c) minimum vapor pressure deficit steadily decreasing between 1 to 4 (Figure 286c); (d) mean dewpoint temperature steadily increasing between -5 to 17 (Figure 28d).

The response curves for the *A. c. lavendulare* in 2011 indicate that the most suitable habitat had the following characteristics: (a) mean temperature steadily decreasing between 11 to 20 degrees Celsius (Figure 29a); (b) precipitation steadily increasing between 1000 to 1500 mm (Figure 29b); (c) minimum vapor pressure deficit steadily decreasing between 1 to 2 (Figure 29c); (d) mean dewpoint temperature slightly increasing (Figure 29d).

The response curves for the *A. c. mobilense* in 2011 indicate that the most suitable habitat had the following characteristics: (a) mean temperature steadily increasing between 11 to 24 degrees Celsius (Figure 30a); (b) precipitation steadily increasing between 500 to 2000 mm (Figure 30b); (c) minimum vapor pressure deficit steadily decreasing between 1 to 3 (Figure 30c); (d) mean dewpoint temperature steadily increasing between 10 to 15 degrees Celsius (Figure 30d).

The response curves for the *A. c. fraseri* in 2011 indicate that the most suitable habitat had the following characteristics: (a) mean temperature steadily increasing between 11 to 24 degrees Celsius (Figure 31a); (b) precipitation slightly decreasing (Figure 31b); (c) minimum vapor pressure slightly increasing (Figure 31c); (d) mean dewpoint temperature steadily increasing between -5 to 17 degrees Celsius (Figure 31d).

The response curves for the *A. canadense* parentals in 2011 indicate that the most suitable habitat had the following characteristics: (a) mean temperature between 19 to 25 degrees Celsius (Figure 32a); (b) precipitation between 400 to 1300 (Figure 32b); (c) minimum vapor pressure deficit steadily decreasing between 0 to 1 (Figure 32c); (d) mean dewpoint temperature between 5 to 18 degrees Celsius (Figure 32d).

The response curves for the *A. canadense* hybrids in 2011 indicate that the most suitable habitat had the following characteristics: (a) mean temperature between 20 to 25 degrees Celsius (Figure 33a); (b) precipitation between 1000 to 1500 (Figure 33b); (c) minimum vapor pressure deficit steadily increasing between 1 to 7 (Figure 33c); (d) mean dewpoint temperature steadily increasing between -5 to 10 degrees Celsius (Figure 33d).

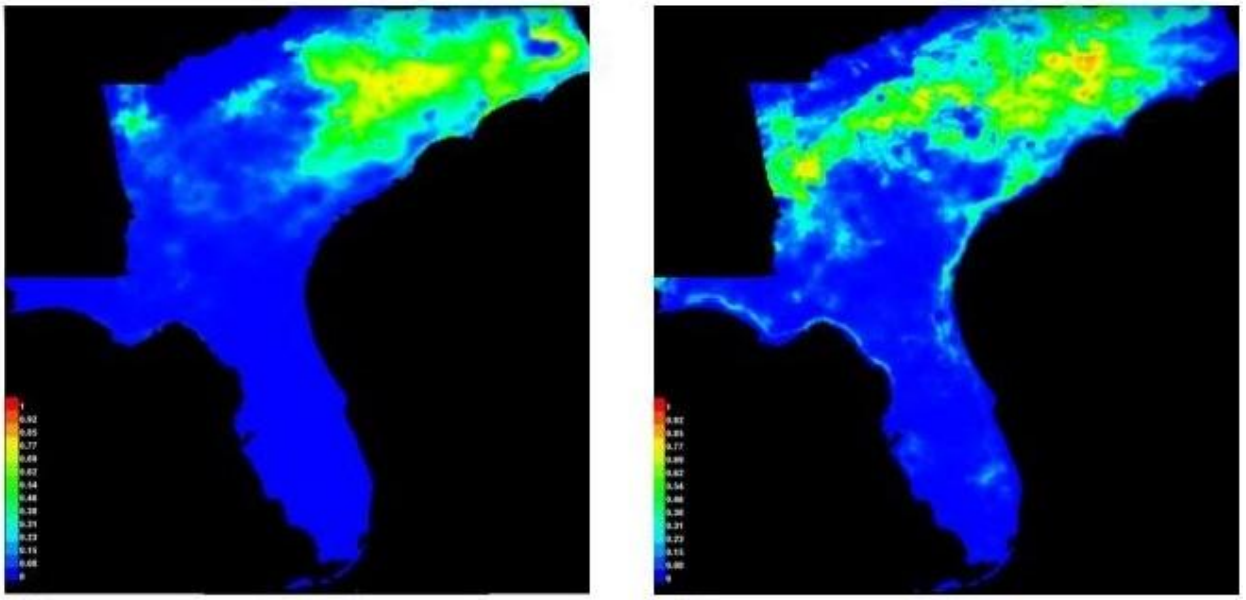
Comparisons of Niche Model Overlap

The Principal Component Analysis (PCA) shows that two components contribute to 96.2% of the total variation among *A. canadense* varieties (PC1= 66.4%, PC2= 29.8%; Figure 34). Precipitation and mean temperature indicates that is equally contributes to PC1 and PC2 (Figure 34). Maximum vapor pressure deficit is more strongly associated with PC1 (Figure 34). The Principal Component Analysis determined that maximum vapor pressure deficit is an important environmental variable used to simulate the suitable habitat for *A. canadense* varieties in 1929 (Figure 34). The PCA also indicated that the varieties are overlapping which indicates that the varieties do not occupy distinct ranges in 1929 (Figure 34).

The Principal Component Analysis shows that two components contribute to 93% of the total variation among *A. canadense* varieties (PC1= 69.8%, PC2= 23.2%; Figure 35). Mean dewpoint temperature and minimum vapor pressure deficit indicates that is equally contributes to PC1 and PC2 (Figure 35). Mean temperature and precipitation are more strongly associated with PC1 (Figure 35). The Principal Component Analysis (PCA) determined that precipitation and mean temperature are important environmental variables used to simulate the suitable habitat for *A. canadense* varieties in 2011 (Figure 35). The PCA also indicated that the varieties are overlapping which indicates that the varieties do not occupy distinct ranges in 2011 (Figure 35).

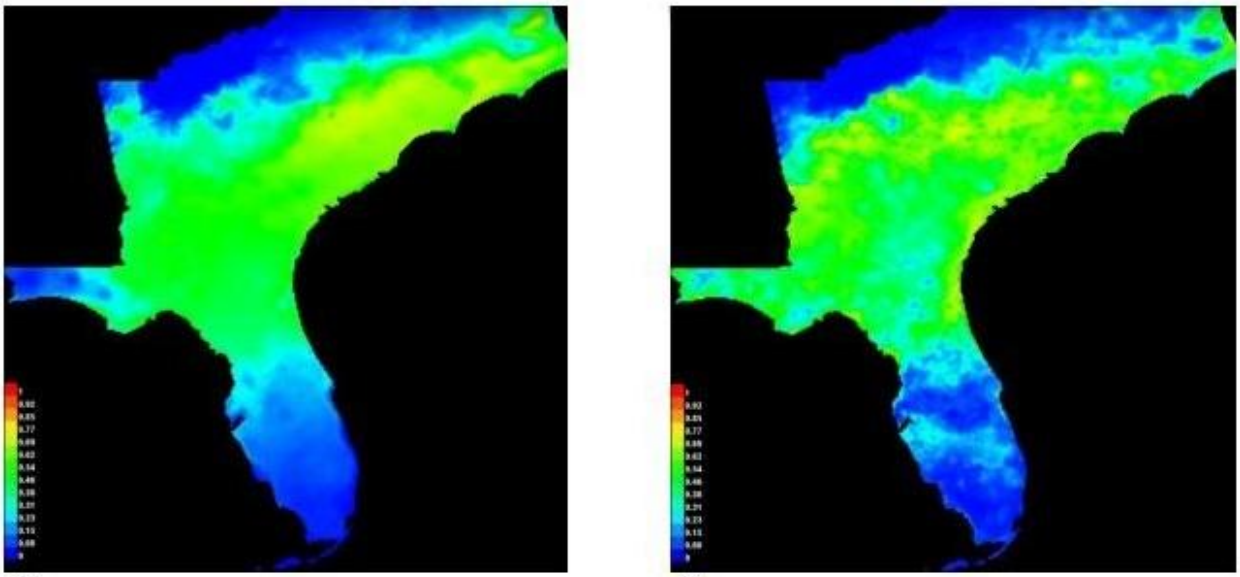
Table 3: Results of *C. graminea* range comparisons. “Combined” means both diploid and tetraploid are included in the same model. An asterisk (*) indicates ranges are significantly different.

range comparison	5% permuted I statistic	real I statistic
1929 - diploid vs tetraploid	0.9482396	0.782628759*
2011 - diploid vs tetraploid	0.9667626	0.800487907*
1929 vs 2011 - diploid	0.894105	0.8623117*
1929 vs 2011 - tetraploid	0.9508259	0.9808046
1929 vs 2011 - combined	0.9300597	0.958536



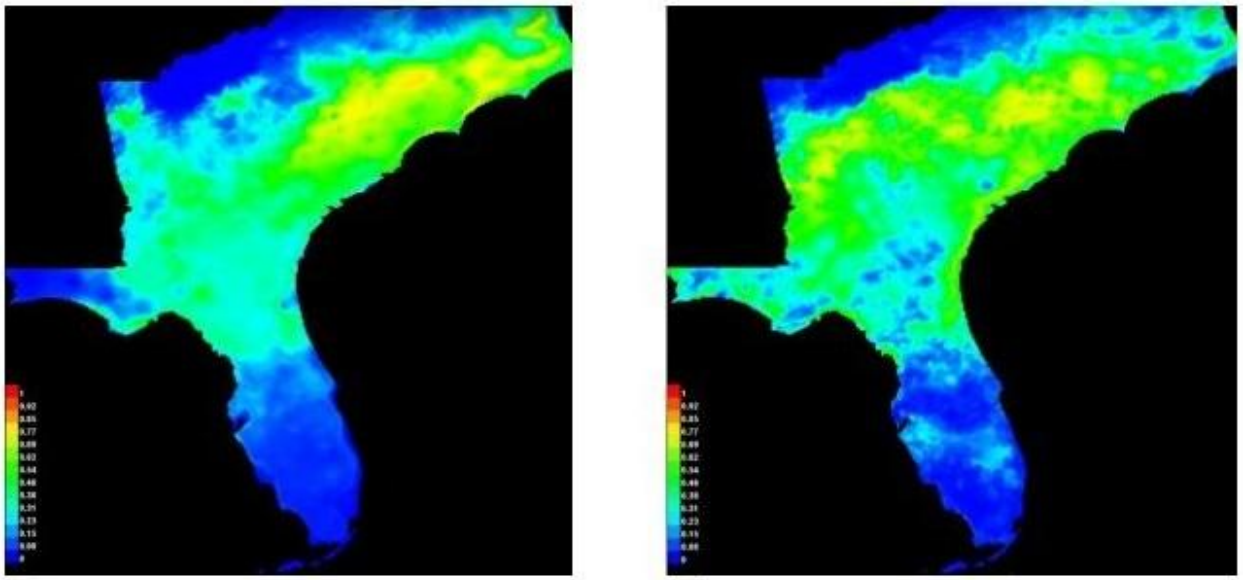
(A) (B)

Figure 3: Models of *C. graminea* diploids in 1929 (A) and 2011 (B).



(A) (B)

Figure 4: Models of *C. graminea* tetraploids in 1929 (A) and 2011 (B).



(A) (B)

Figure 5: Combined Models of *C. graminea* diploids and tetraploids in 1929 (A) and 2011 (B).

Table 4: Test gain for *C. graminea* cytotypes for 1929.

Taxon	Test AUC	Full model	Only precipitation	Only mean temperature	Only minimum vapor pressure deficit
diploids	0.9709	1.9357	0.8776	1.1771	0.6953
tetraploids	0.8016	0.5605	0.1428	0.4950	0.1239
both cytotypes	0.8158	0.6739	0.3407	0.4962	0.1126

Table 5: Test gain for *C. graminea* cytotypes for 2011.

Taxon	Test AUC	Full model	Only precipitation	Only mean temperature	Only maximum vapor pressure deficit	Only minimum vapor pressure deficit
diploids	0.9728	1.8331	0.4065	0.9421	1.1362	0.1384
tetraploids	0.7532	0.4756	0.1602	0.4336	−0.0401	0.0264
both cytotypes	0.8265	0.6891	0.2534	0.4426	0.3358	−0.0009

Figure 6: Response curves for *C. graminea* diploids for 1929.

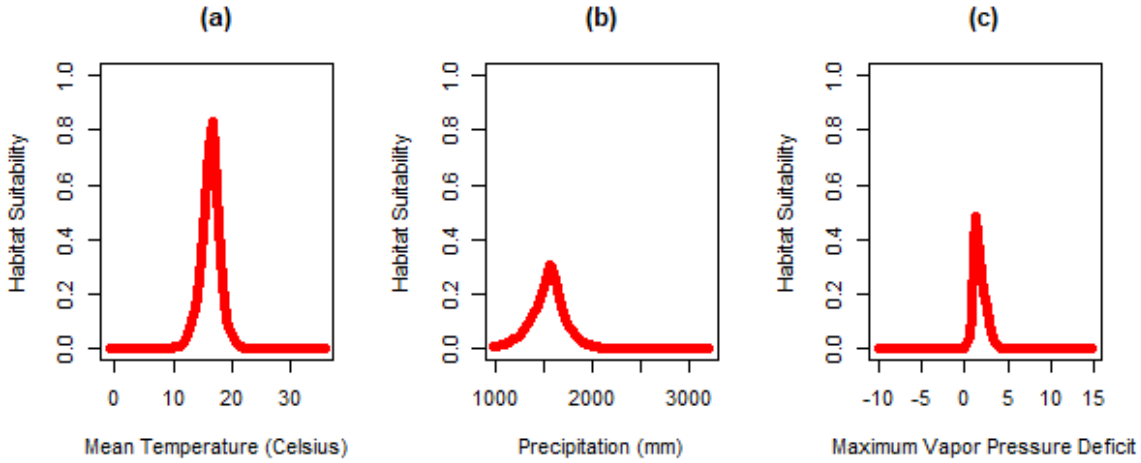


Figure 7: Response curves for *C. graminea* tetraploids for 1929.

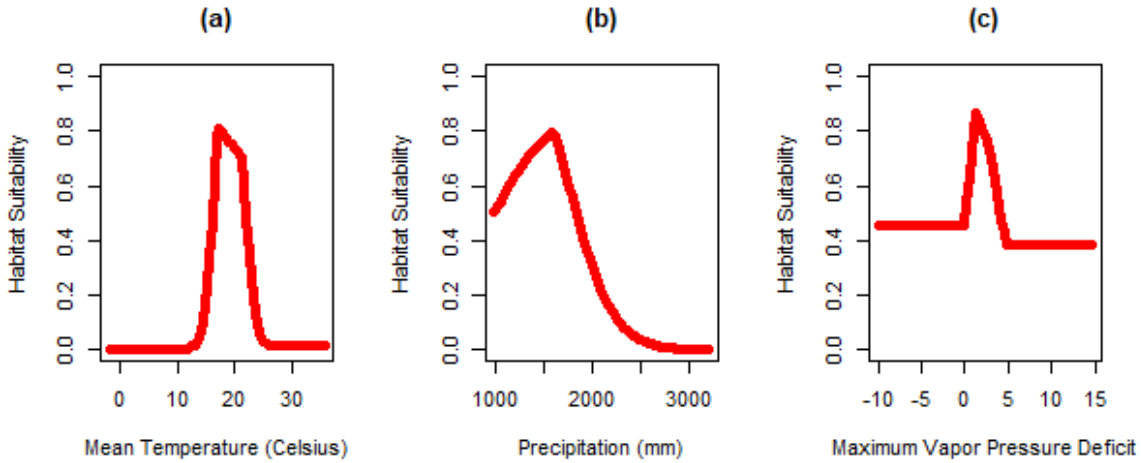


Figure 8: Response curves for both cytotypes of *C. graminea* for 1929.

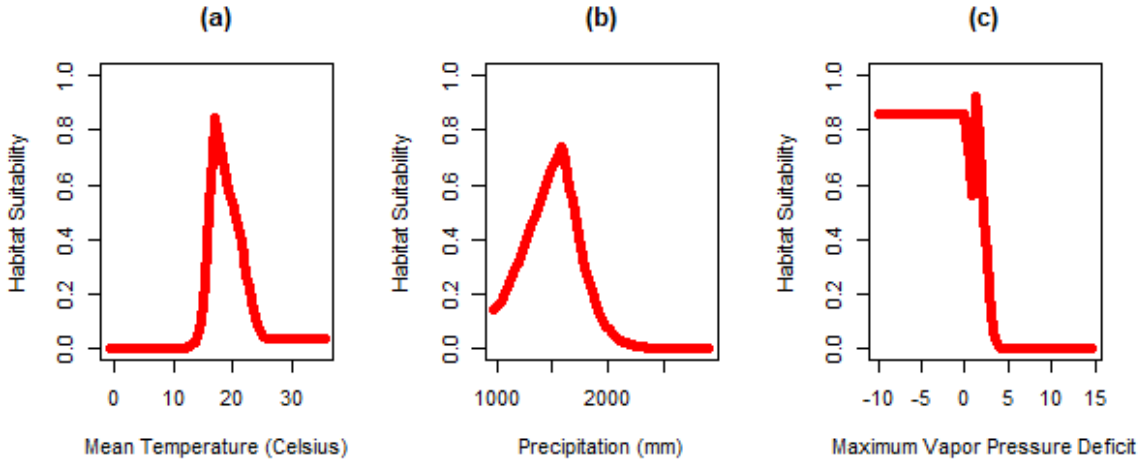


Figure 9: Response curves for *C. graminea* diploids for 2011.

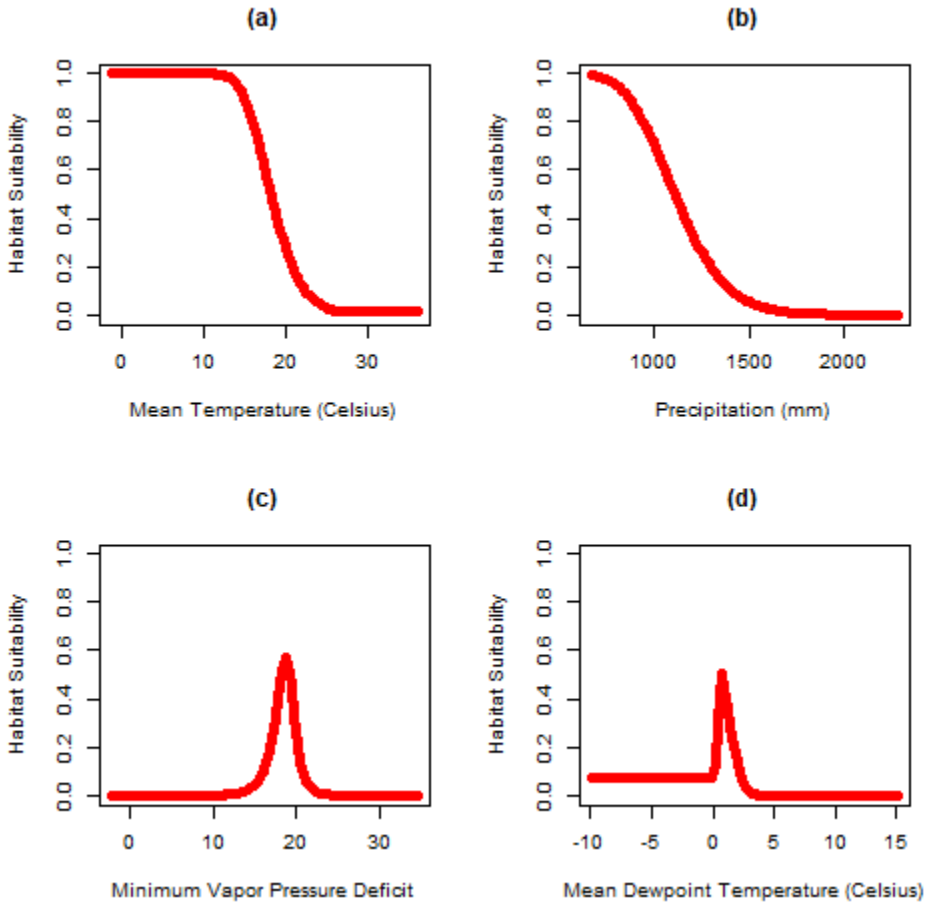


Figure 10: Response curves for *C. graminea* tetraploids for 2011.

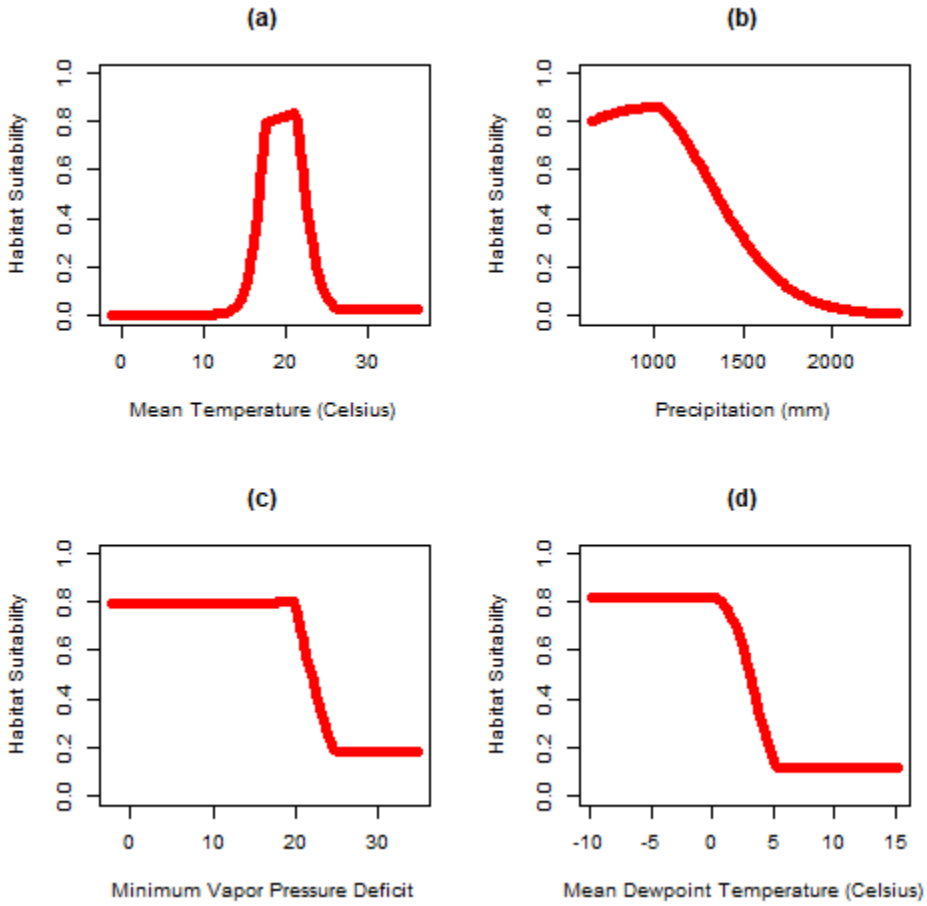
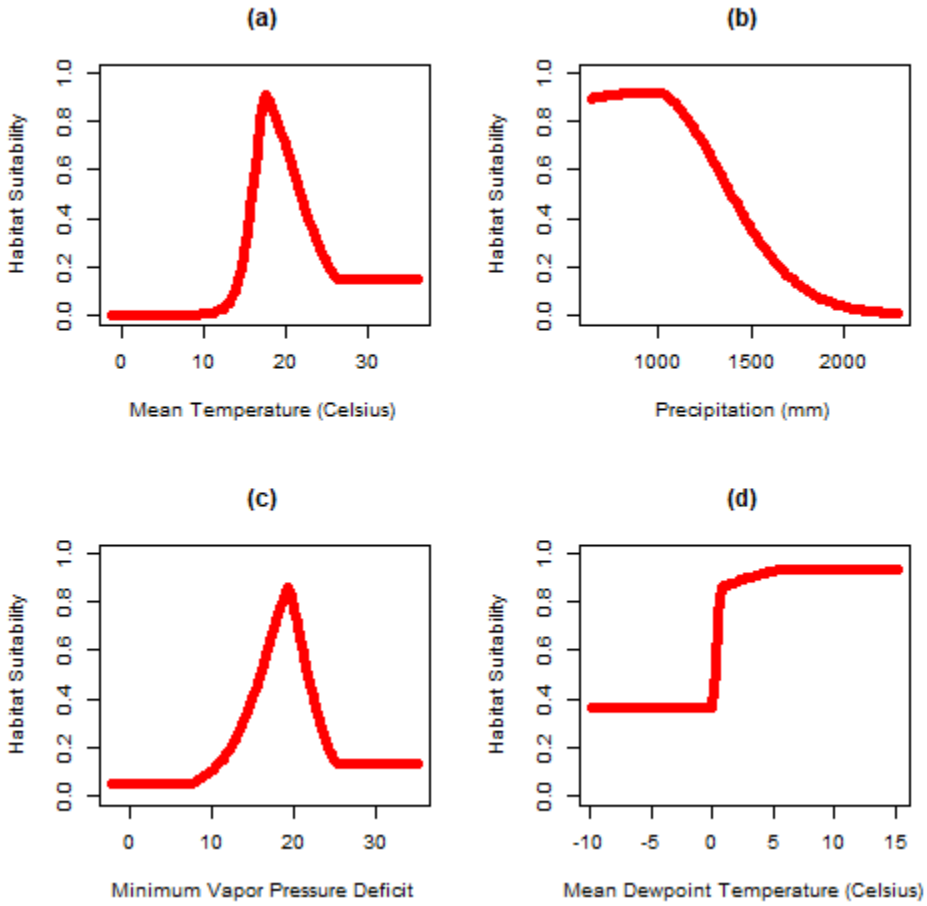


Figure 11: Response curves for both cytotypes of *C. graminea* for 2011.



Comparisons of Niche Model Overlap

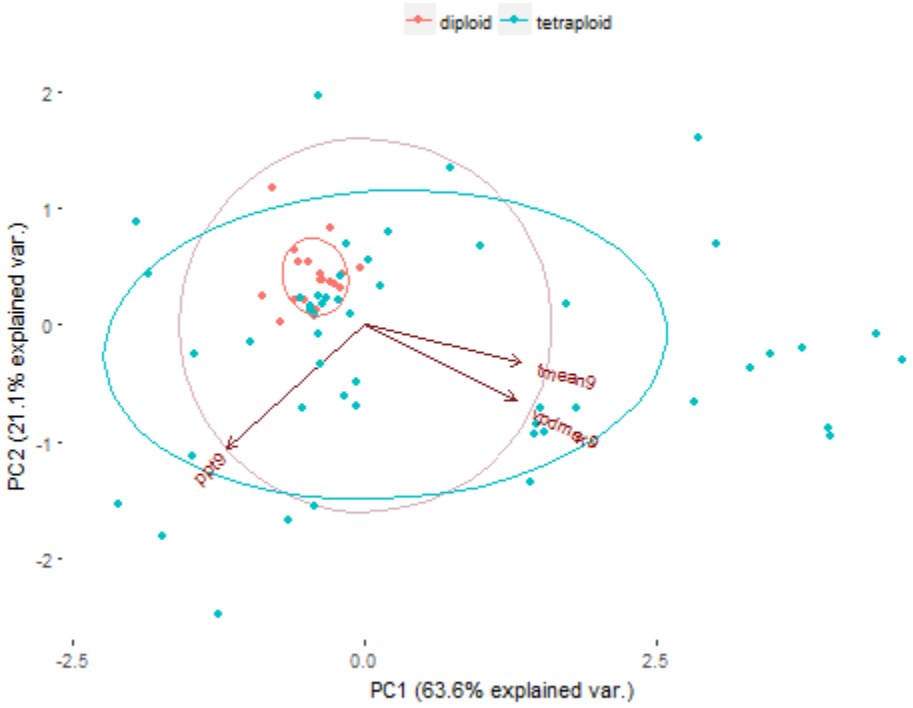


Figure 12: A plot of the Principal Components Analysis (PCA) for *C. graminea* cytotypes for 1929. The purple circle accompanies the display of variables and is not included in the analysis.

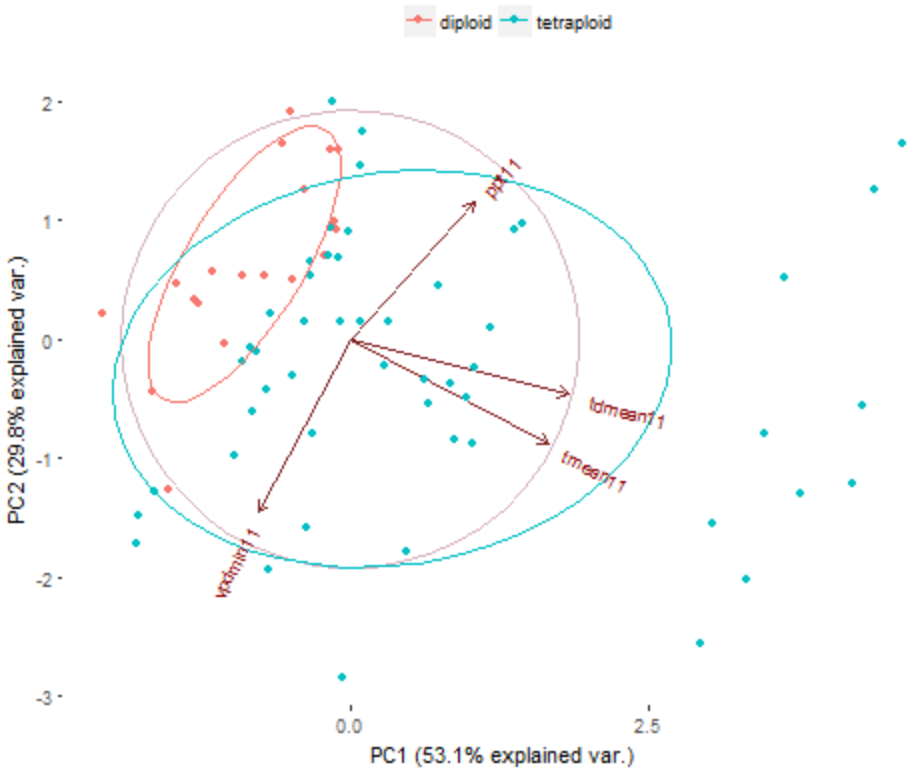


Figure 13: A plot of the Principal Components Analysis (PCA) for *C. graminea* cytotypes for 2011. The purple circle accompanies the display of variables and is not included in the analysis.

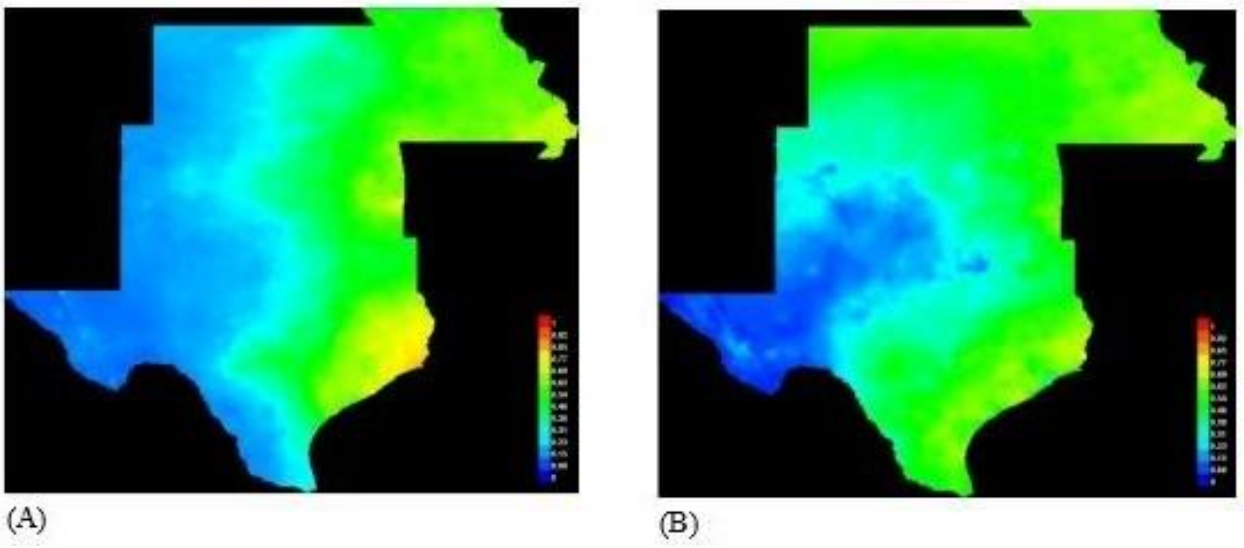
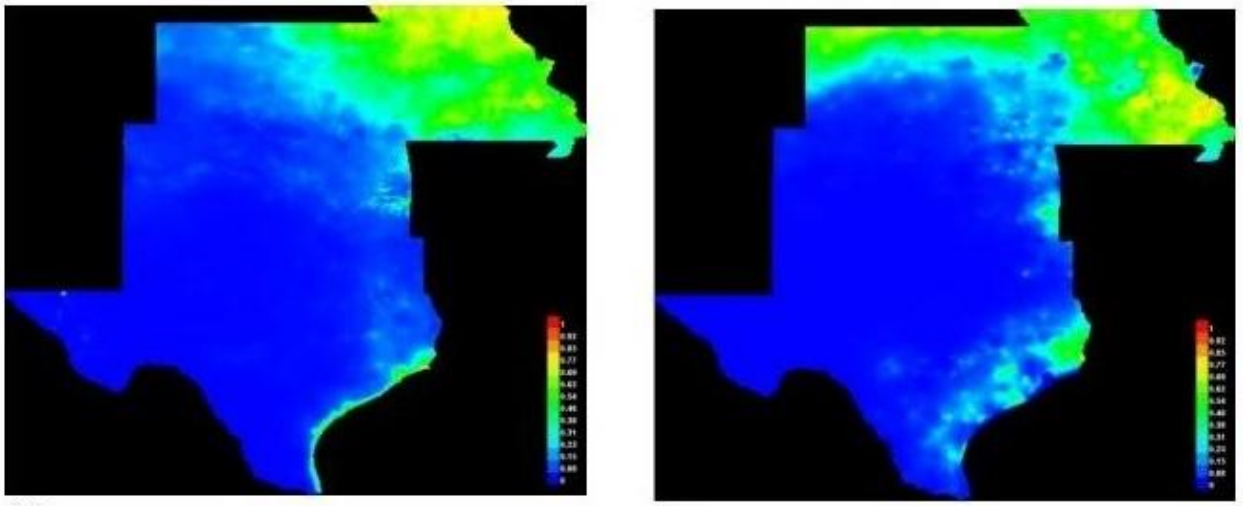
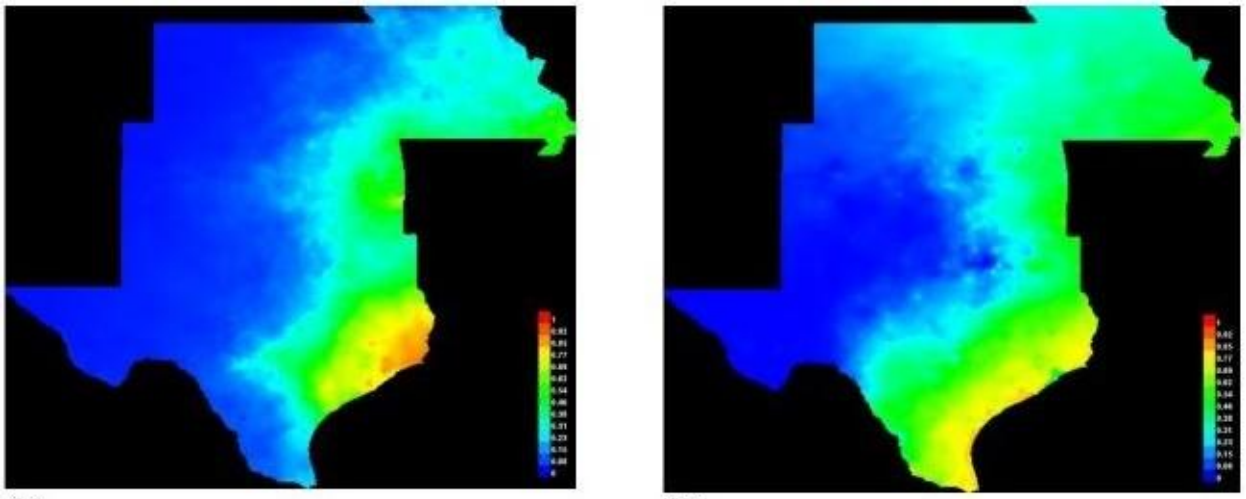


Figure 14: Models of *A. c. canadense* in 1929 (A) and 2011 (B).



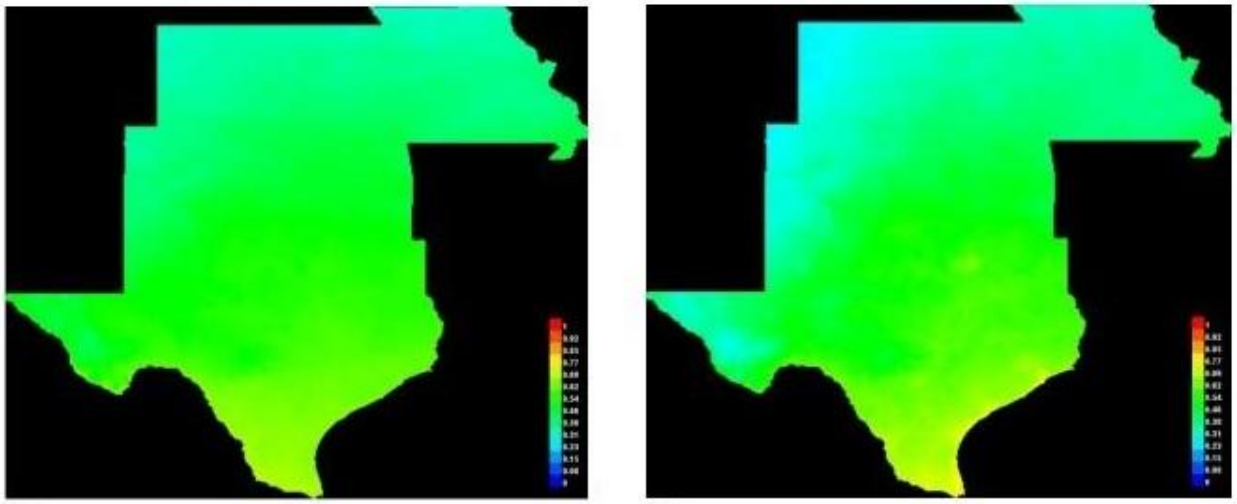
(A) (B)

Figure 15: Models of *A. c. lavendulare* in 1929 (A) and 2011 (B).



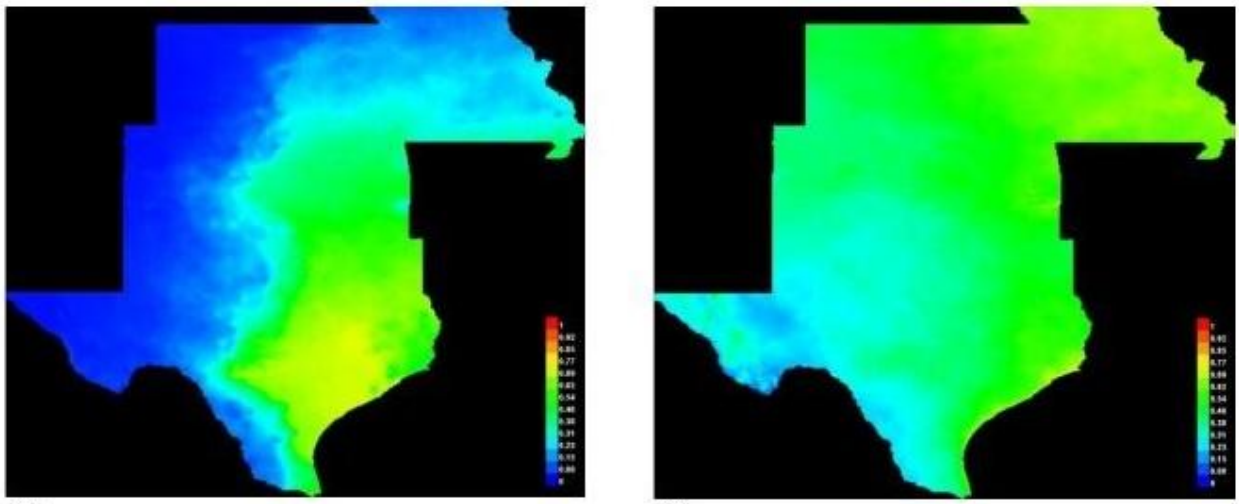
(A) (B)

Figure 16: Models of *A. c. mobilense* in 1929 (A) and 2011 (B).



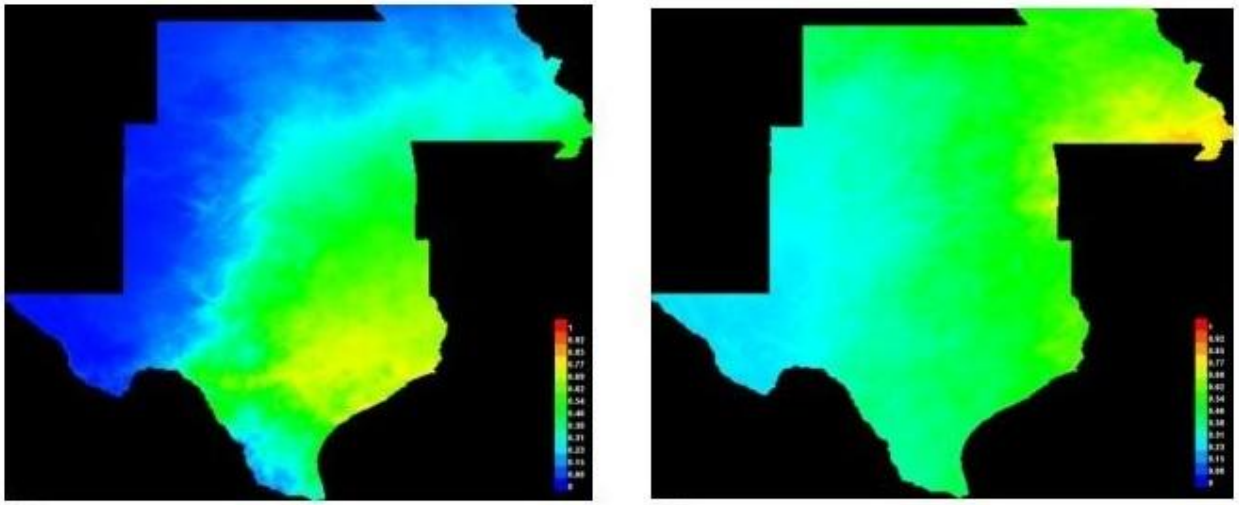
(A) (B)

Figure 17: Models of *A. c. fraseri* in 1929 (A) and 2011 (B).



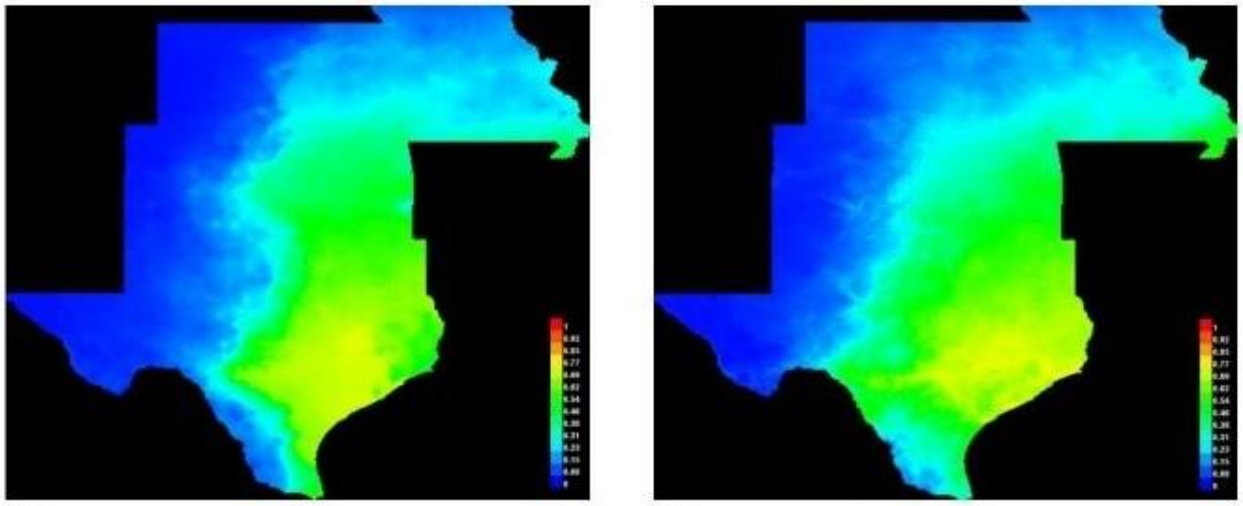
(A) (B)

Figure 18: Models of *A. canadense* parentals in 1929 (A) and hybrids in 1929 (B).



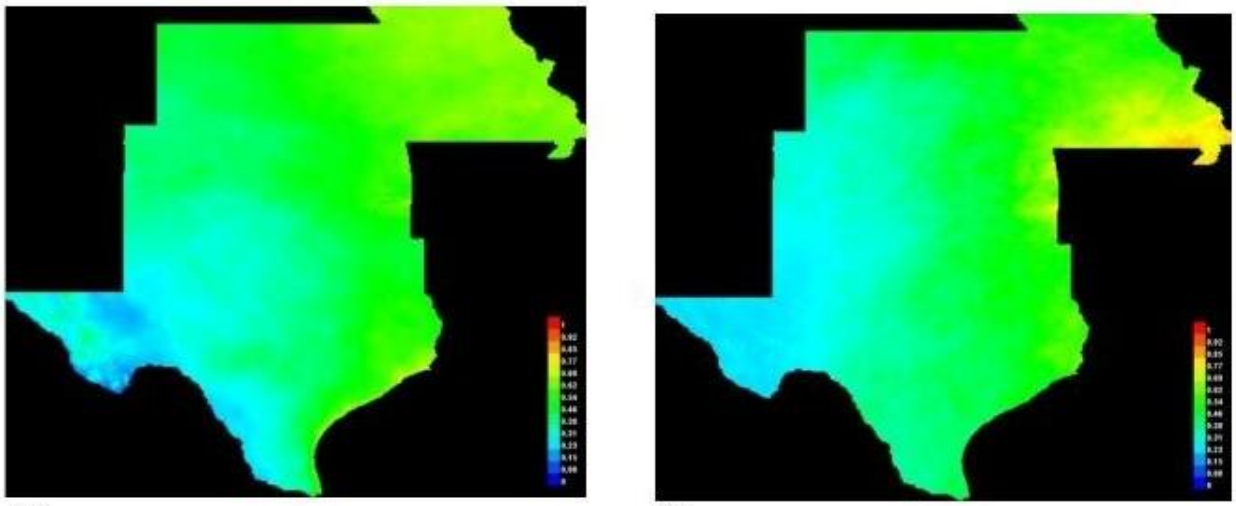
(A) (B)

Figure 19: Models of *A. canadense* parentals in 2011 (A) and hybrids in 2011 (B).



(A) (B)

Figure 20: Models of *A. canadense* parentals in 1929 (A) and 2011 (B).



(A) (B)

Figure 21: Models of *A. canadense* hybrids in 1929 (A) and 2011 (B).

Table 6: Results of *A. canadense* range comparisons. "Combined" means all six varieties are included in the same model. var. *A. c. ecristatum* not tested because too few occurrence points were represented in the dataset.

range comparison	5% permuted I statistic	real I statistic
1929 vs 2011 - canadense	0.8708627	0.9706688
1929 vs 2011 - lavendulare	0.7734544	0.922259
1929 vs 2011 - mobilense	0.8566591	0.9454854
1929 vs 2011 - fraseri	0.8969706	0.9982555
1929 vs 2011 - parentals	0.9273717	0.9893907
1929 vs 2011 - hybrids	0.9447226	0.9952878
1929 - parentals vs hybrids	0.9119269	0.917279519
2011 - parentals vs hybrids	0.931404	0.941136127

Table 7: Test gain for *A. canadense* varieties for 1929.

Taxon	Test AUC	Full model	Only precipitation	Only mean temperature	Only maximum vapor pressure deficit
canadense	0.7734	0.4243	0.4565	−0.0047	0.1205
lavendulare	0.8882	0.9737	0.2647	0.6993	1.3295
mobilense	0.9038	0.9241	0.9246	0.2176	0.1134
fraseri	0.7051	0.0231	−0.0004	0.0238	−0.0548
parentals	0.7777	0.5761	0.5291	0.2633	0.0840
hybrids	0.7858	0.5452	0.4408	0.4004	−0.2026

Table 8: Test gain for *A. canadense* varieties for 2011.

Taxon	Test AUC	Full model	Only precipitation	Only mean dewpoint temperature	Only mean temperature	Only minimum vapor pressure deficit
canadense	0.7130	0.2417	0.1586	0.0646	−0.0198	0.3032
lavendulare	0.9258	1.4006	0.7270	−0.0097	0.7122	1.3664
mobilense	0.8585	0.7063	0.0428	0.6155	0.1532	0.3833
fraseri	0.6809	0.0062	−0.0254	0.1067	0.0340	−0.0245
parentals	0.7637	0.4749	0.2960	0.3436	0.3563	0.0176
hybrids	0.6750	0.1725	0.2912	0.0835	−0.0597	0.1171

Figure 22: Response curves for *A. c. canadense* for 1929.

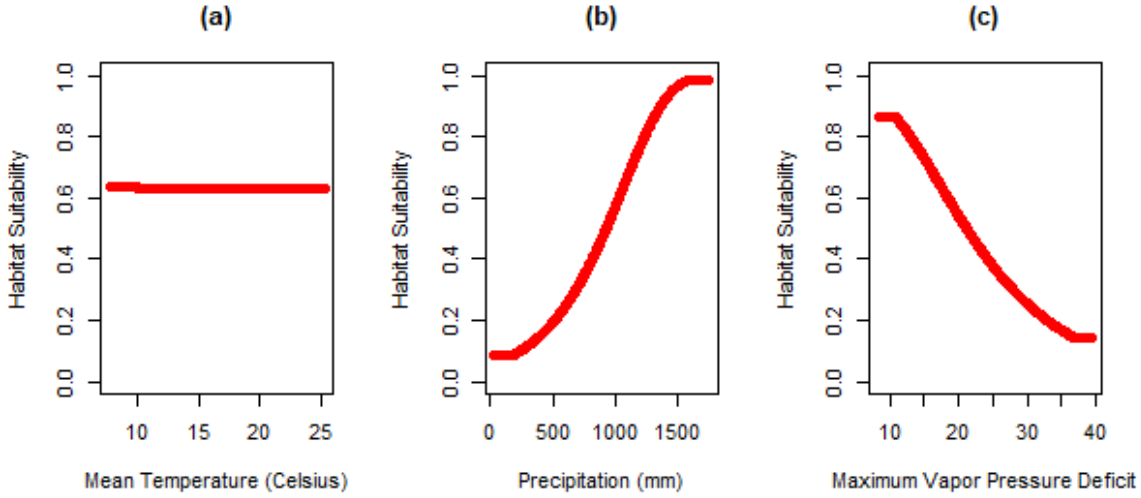


Figure 23: Response curves for *A. c. lavendulare* for 1929.

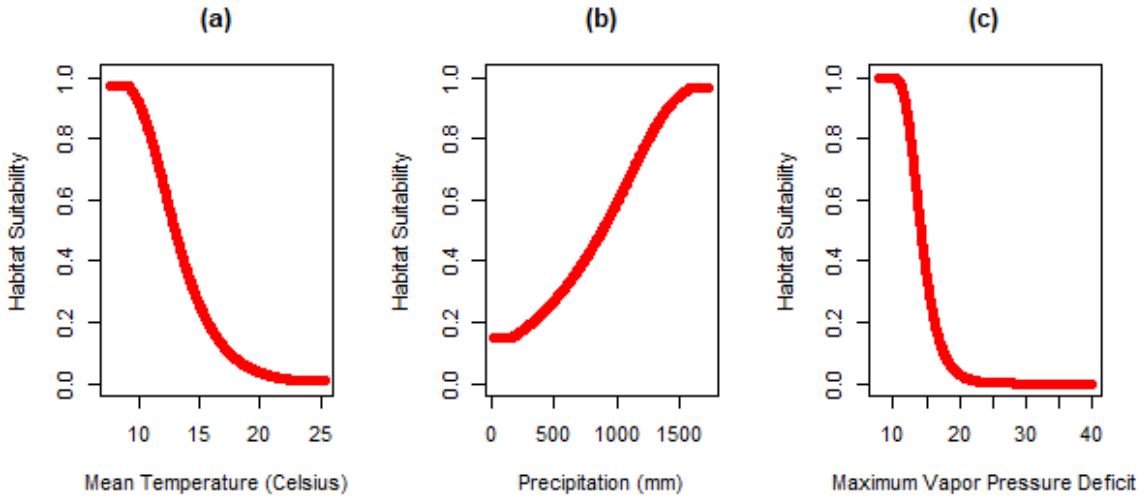


Figure 24: Response curves for *A. c. mobilense* for 1929.

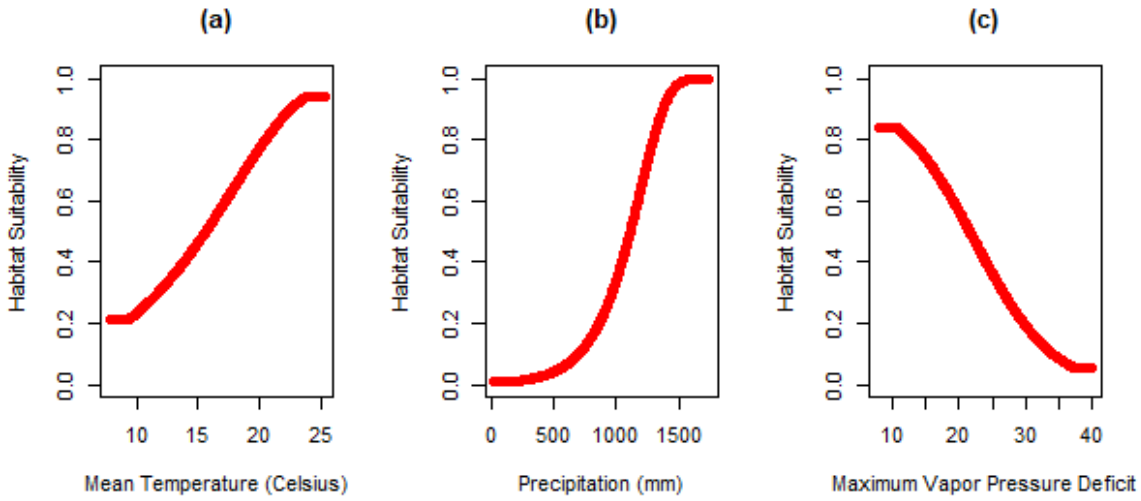


Figure 25: Response curves for *A. c. fraseri* for 1929.

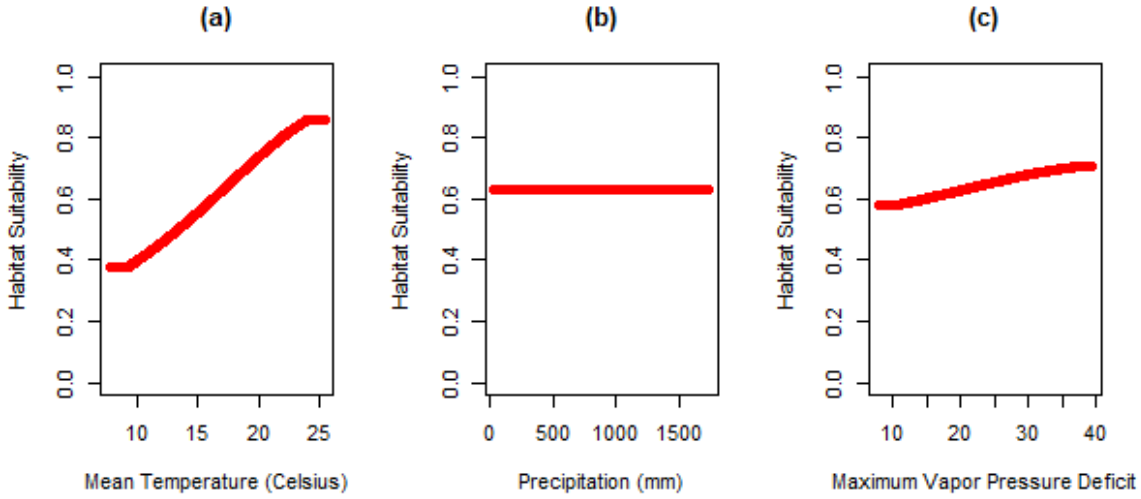


Figure 26: Response curves for *A. canadense* parentals for 1929.

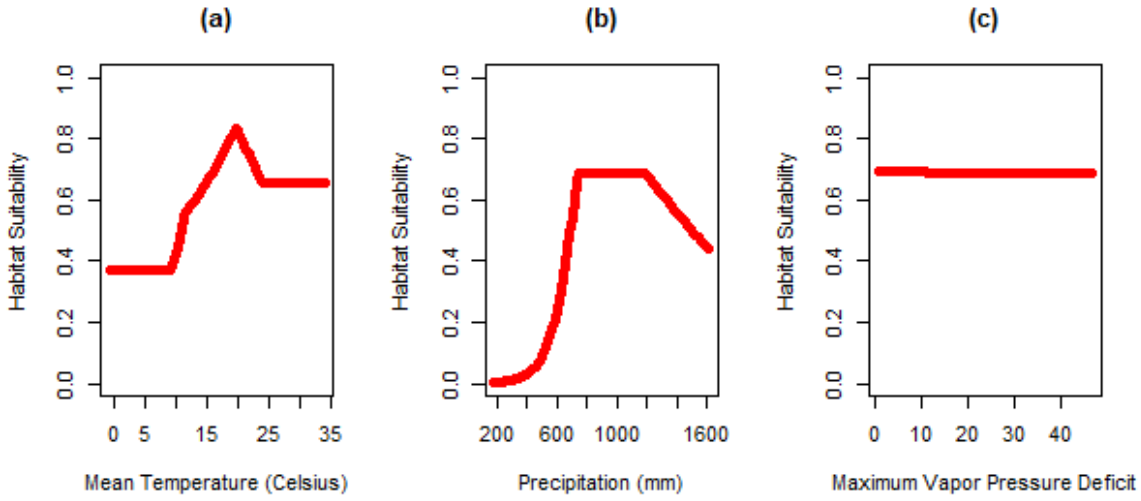


Figure 27: Response curves for *A. canadense* hybrids for 1929.

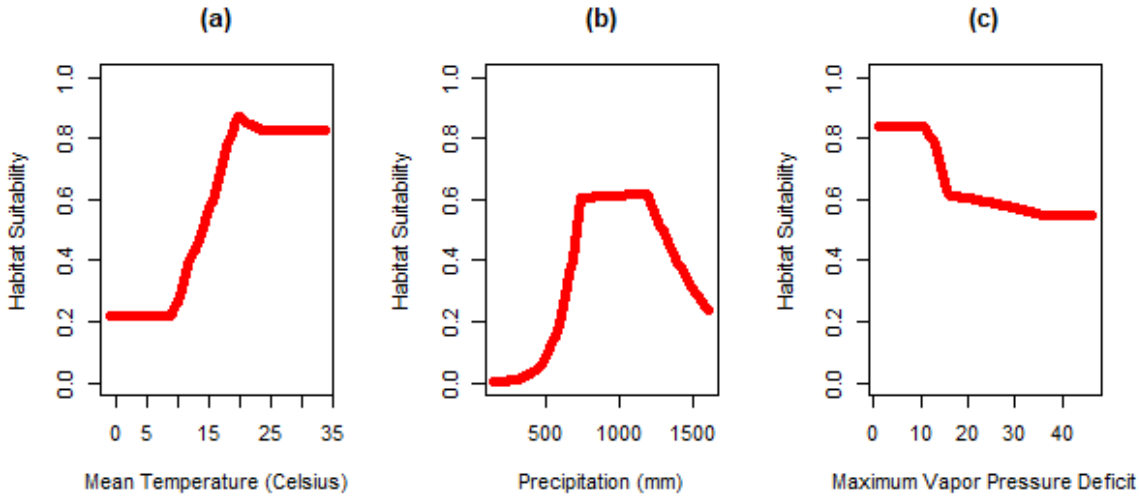


Figure 28: Response curves for *A. c. canadense* for 2011.

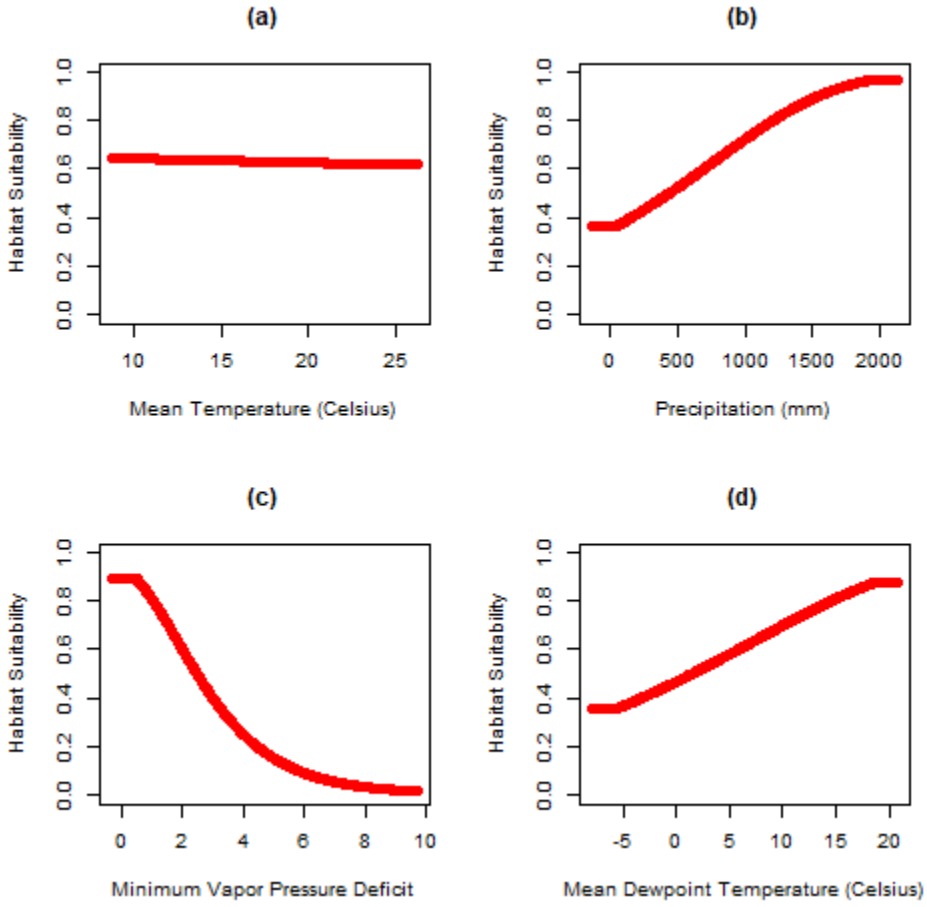


Figure 29: Response curves for *A. c. lavendulare* for 2011.

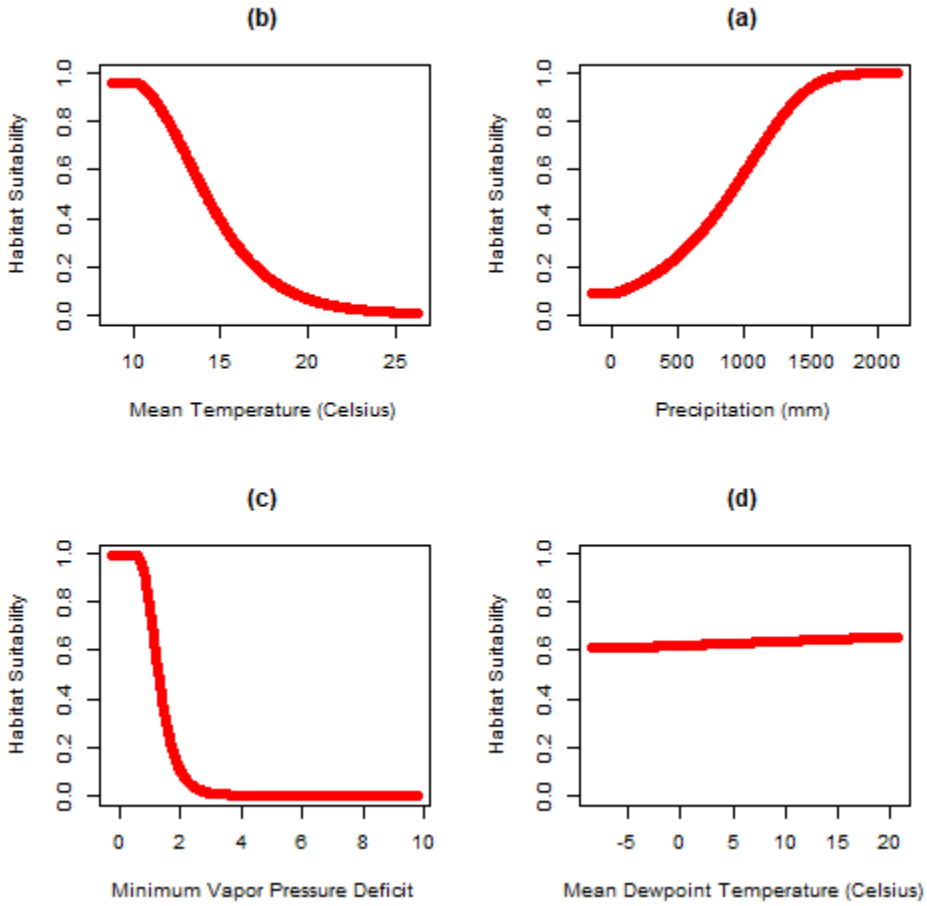


Figure 30: Response curves for *A. c. mobilense* for 2011.

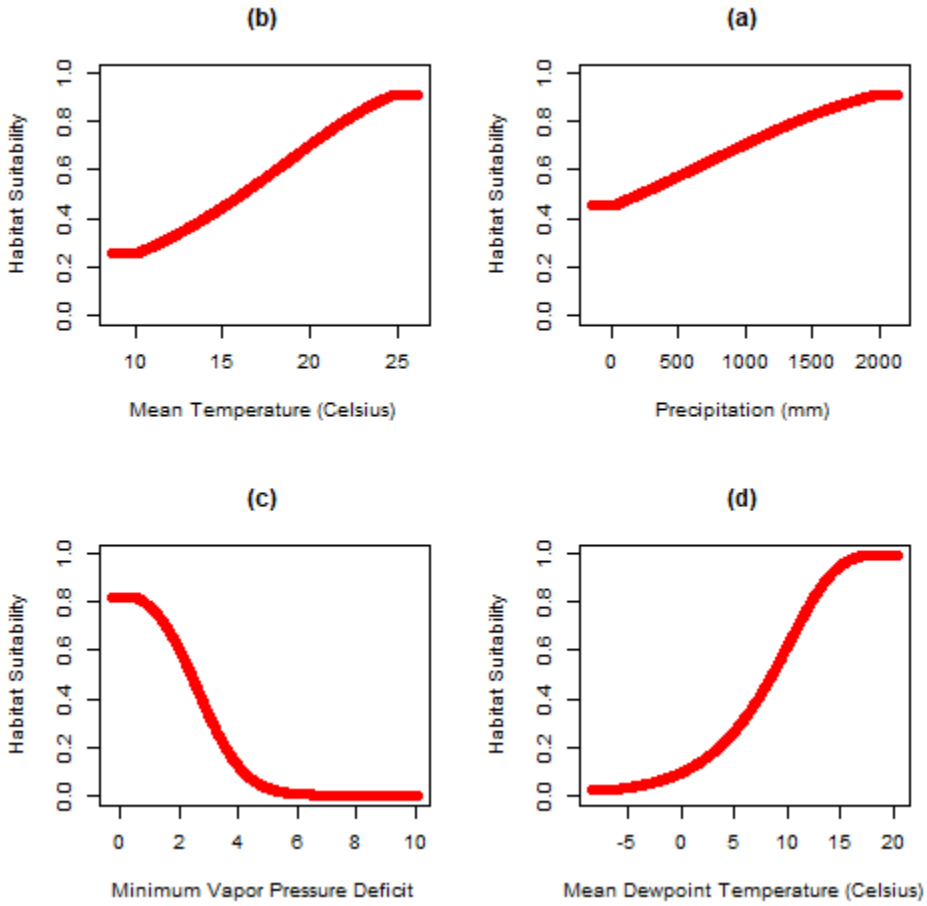


Figure 31: Response curves for *A. c. fraseri* for 2011.

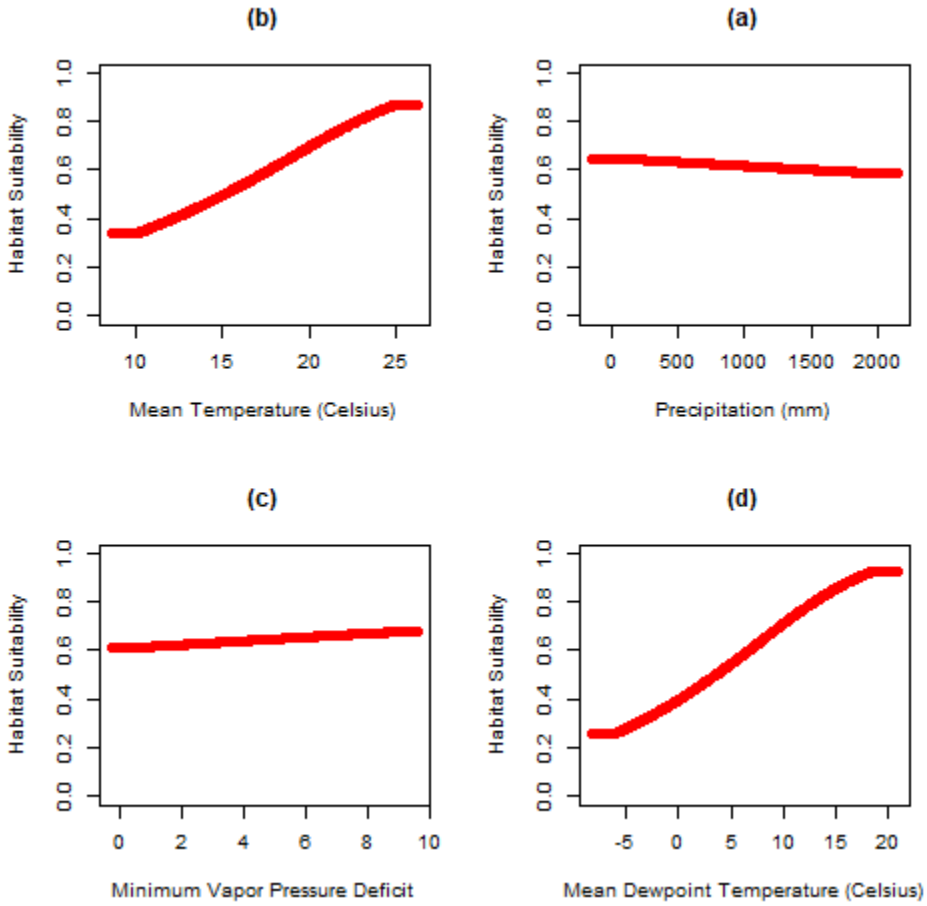


Figure 32: Response curves for *A. canadense* parentals for 2011.

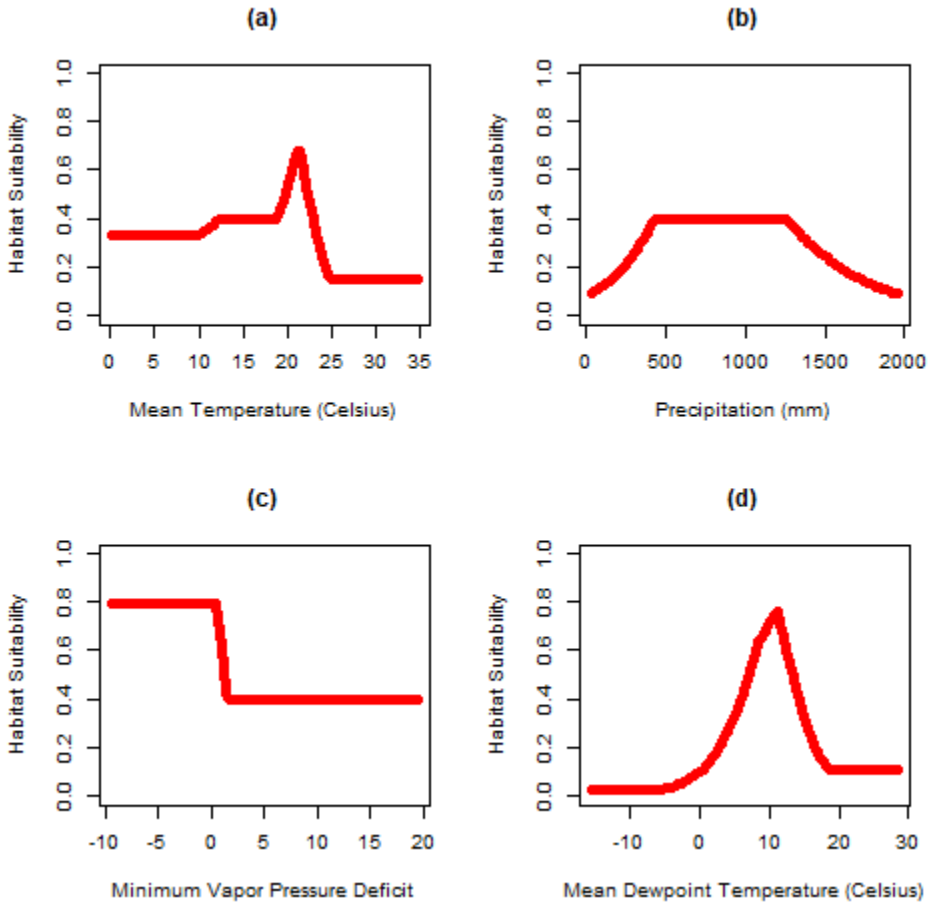
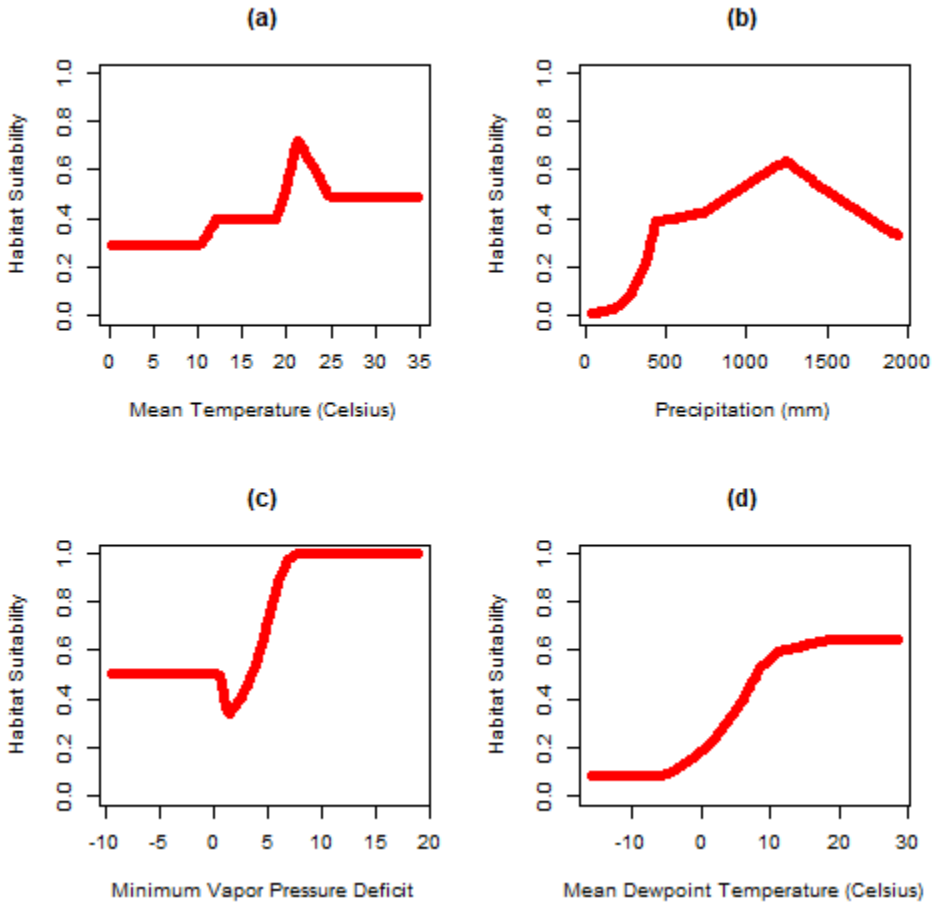


Figure 33: Response curves for *A. canadense* hybrids for 2011.



Comparisons of Niche Model Overlap

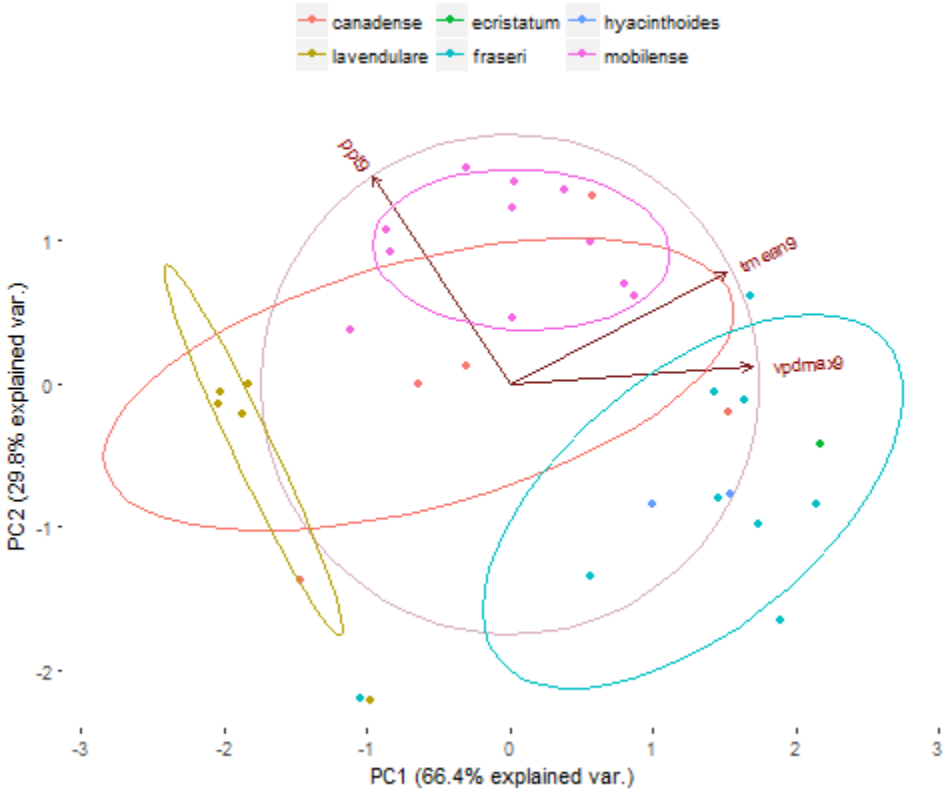


Figure 34: A plot of the Principal Components Analysis (PCA) for *A. canadense* varieties for 1929. The purple circle accompanies the display of variables and is not included in the analysis.

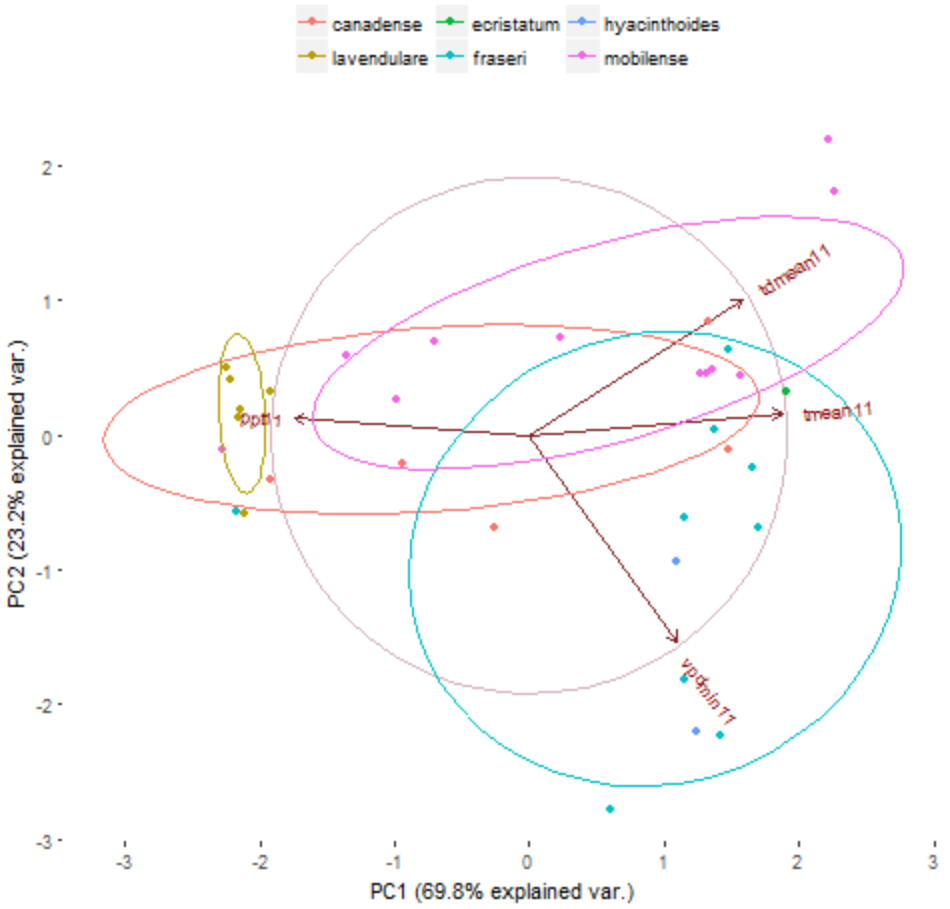


Figure 35: A plot of the Principal Components Analysis (PCA) for *A. canadense* varieties for 2011. The purple circle accompanies the display of variables and is not included in the analysis.

Chapter Four
Discussion

This project applied niche modeling to investigate the effects of polyploidy and hybridization on the distribution of *C. graminea* and *A. canadense* over time. Below I highlight major findings and suggestions for future research on this topic.

I expected that the *C. graminea* tetraploids range would expand over time and the *C. graminea* diploids range would remain constant. The increase in niche overlap between the *C. graminea* cytotypes could indicate the beginning of sympatric speciation (Table 3). Chromosome doubling immediately ceases gene flow by causing the cytotypes to have different flowering times and dissimilar flower morphology to attract different pollinators (Baack and Stanton, 2005). Any cross-pollination with diploids will create a triploid offspring that has low fitness; to avoid this reproductive act, selfing, self-fertilization, is often utilized by newly formed tetraploids (Baack and Stanton, 2005). The diploid and tetraploid ranges are significantly different from one another, which show these cytotypes are exhibiting competitive exclusion and environmental adaptation (Ramsey and Ramsey, 2014). Polyploids have more genetic material and biodiversity, therefore those cytotypes should exhibit an elevated ecological tolerance and be able to maintain a broader, distinct niche compared to their diploid progenitors (Theodoridis et al., 2013). The diploids and tetraploids occupy distinct ranges despite an increase in niche overlap (Table 3). The cytotypes demonstrate competitive exclusion, reproductive isolation that are seen in polyploids in the field; along with a substantial phenotypic divergence one could suggest sympatric speciation in the *C. graminea* system (Ramsey and Ramsey, 2014).

I expected that the variety *A. c. lavendulare*, which has a diploid and tetraploid population, range would expand over time. I expected that the var. *A. c. lavendulare* range would not remain consistent. The *A. c. lavendulare* range expanded which is not unexpected since it is both a hybrid and a polyploid. Higher biodiversity is obtained from being a polyploid; hybrids show this same trend when compared to their parental species (Soltis et al., 2014; Husband et al., 2013). The ability to adapt and expand the niche according to favorable environmental conditions through time is a crucial element for plants (Theodoridis et al., 2013; Johnson and Ashman, 2014; Thompson et al., 2014).

I expected that the parental varieties *A. c. mobilense* and *A. c. fraseri* would have ranges that remain consistent whereas the hybrid varieties *A. c. lavendulare*, *A. c. ecristatum*, and *A. c. hyacinthoides* would expand. As expected, the *A. canadense* hybrids range was larger than the *A. canadense* parentals range. When comparing the parentals in 1929 to 2011, the range has expanded whereas the hybrid range has contracted. Interestingly, counterintuitive cases like this have occurred before such as the *Centaurea maculosa* diploid outperforming its tetraploid in Europe (Treier et al., 2009). A species distribution is determined by the fluctuating stability in location of available, suitable, required environmental conditions (Pulliam, 2000). The retraction or expansion of the taxon's range could be a reaction provoked by a lack or abundance of habitat availability and stability via the level of interspecific competition (Pulliam, 2000). This could be direct from predation by herbivores, indirect by competition for sunlight by nearby vegetation, or limited dispersal (Pulliam, 2000). In order to evaluate if the trends revealed through niche modeling hold true, more environmental layers and observations over a longer time span need to be incorporated in these models.

The assumption of this work was that polyploids have more genetic material and biodiversity, therefore the polyploids should exhibit an elevated ecological tolerance and be able to maintain a broader, distinct range compared to their diploid progenitors. But there was an unexpected expansion in the diploid range for one taxon (Figure 3). This expansion may be due to differences in floral structure or pollinators between the cytotypes (Levin, 2003). Alterations to the environment by disturbance could cause ranges to decrease (Levin, 2003). It would be interesting to see how the models change when soil layers and rivers are included. The proximity to water and/ or arid soils could impact responsiveness to climate variables like vapor pressure deficit and precipitation. The selection of the environmental layers is crucial to the accurate assessment of the suitable habitat of the taxa. The layers used for building niche models were restricted to those available on the PRISM website. High correlation between the available PRISM layers restricted the complexity of the models. Soil layers were not included due to difficulty in acquiring contiguous layers for entire states.

In addition to selection of climate layers, effectiveness of modeling is also reliant on robustness of species occurrence data. Reliable predictive models cannot be built for species with less than five available occurrence points for each taxon. It is difficult to assess niche models for narrowly defined taxa with few available occurrence points, such as rare and elusive species, that have a limited spatial distribution pattern, few occurrence points, and some tend to be habitat specialists. They are difficult to track and have a limited number of sites in which they are known to occur. Their small sample sizes make statistical analyses difficult and decreases predictability of the niche model, when compared with other models that incorporated more occurrence points. The accuracy of the niche model increases with an increased sample size until it reaches its maximal

accuracy potential and plateaus. The sample size and maximum accuracy potential where the asymptote will be reached depends on the study area and species of interest, the spatial resolution and quality of the environmental layers and occurrence data, and the modeling method employed (Hernandez, 2006).

This thesis focused on the consequences of autopolyploidy in plant systems. In the 1980s and 1990s, the ecological significance of autopolyploidy began to be accepted and explored by population biologists in plant systems (Ramsey and Ramsey, 2014). Less attention has been place on allopolyploid complexes and their distribution due to underdeveloped tools by molecular systematists and confusion concerning phylogenetic relationships (Ramsey and Ramsey, 2014). Comparing autopolyploidy and allopolyploidy is an important step to understanding the overall influence of genome duplication in plant distribution or evolution in general. The additional genetic material of polyploids enable them to colonize and adapt to new environments over time (Levin, 2003). Polyploid species, whether a hybrid or autotetraploid, are responsive to their environment due to their excess genetic material. The manner in which they or their progenitors respond depends on their interactions with numerous environmental and biotic factors. By incorporating biotic factors like herbivore resistance and abiotic factors such as size, soil type, drought tolerance, and reproductive systems, niche modeling could help us gain a better understanding of these and other plant systems.

Chapter Five
Bibliography

1. Giles Jr, N. H. Studies of Natural Populations of *Cuthbertia graminea* in the Carolinas. *J. Journal of the Mitchell Society* (1943).

2. Giles Jr, N. H. Autopolyploidy and geographical distribution in *Cuthbertia graminea* Small. *American Journal of Botany* 29, 637–645 (1942).

3. Glennon, K. L., Ritchie, M. E. & Segraves, K. Evidence for shared broad-scale climatic niches of diploid and polyploid plants. *Ecology Letters* 17, 574–582 (2014).

4. Johnson, A. L., Govindarajulu, R. & Ashman, T. L. Bioclimatic evaluation of geographical range in Fragaria (Rosaceae): Consequences of variation in breeding system, ploidy and species age. *Botanical Journal of the Linnean Society* 176, 99–114 (2014).

5. Ramsey, J. & Ramsey, T. S. Ecological studies of polyploidy in the 100 years following its discovery. *Philosophical Transactions of the Royal Society B* 369**,** 1–76 (2014).

6. Soltis, D. E., Visger, C. J. & Soltis, P. S. The polyploidy revolution then...and now: Stebbins revisited. *American Journal of Botany* 101, 1057–1078 (2014).

7. Theodoridis, S., Randin, C., Broennimann, O., Patsiou, T. & Conti, E. Divergent and narrower climatic niches characterize polyploid species of european primroses in primula sect. aleuritia. *Journal of Biogeography* 40**,** 1278–1289 (2013).

8. Thompson, K. A., Husband, B. C. & Maherali, H. Climatic niche differences between diploid and tetraploid cytotypes of *Chamerion angustifolium* (Onagraceae). *American Journal of Botany* 101, 1868–1875 (2014).

9. Treier, U. A., Broennimann, O., Normand, S., Guisan, A., Schaffner, U., Steinger, T., & Muller-Scharer, H. Shift in cytotype frequency and niche space in the invasive plant *Centaurea maculosa*. *Ecology* 90, 1366–1377 (2009).

10. Wheeler, E. J. Phylogenetic and phylogenomic studies of wild onions (*Allium*, Amaryllidaceae) at three taxonomic scales. 1–105 (2011).

11. Hijmans, R. J. (2015). raster: Geographic Data Analysis and Modeling. R package version 2,4-15. http://CRAN.R-project.org/package=raster

12. Hadley, W. & Francois, R. (2015). dplyr: A Grammar of Data Manipulation. R package version 0.4.3. http://CRAN.R-project.org/package=dplyr

13. R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna, Austria. URL http:/www.R-project.org/.

14. Thuiller, W., Georges, D., & Engler, R. (2014). biomod2: Ensemble platform for species distribution modeling. R package version 3.1-64. http://CRAN.R-project.org/package=biomod2

15. Hijmans R. J., Steven Phillips, Leathwick, J., & Elith, J. (2015). dismo: Species Distribution Modeling. R package version 1.0-12. http://CRAN.R-project.org/package=dismo

16. RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

17. Jiao, Y. et al. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100 (2011).

18. Magadum, S., Banerjee, U., Murugan, P., Gangapur, D. & Ravikesavan, R. Gene duplication as a major force in evolution. *Journal of Genetics* 92, 155–161 (2013).

19. Mühlhausen, S. & Kollmar, M. Whole genome duplication events in plant evolution reconstructed and predicted using myosin motor proteins. *BMC Evolutionary Biology* 13, 202 (2013).

20. Soltis, D. E., Albert, V. A., Leebens-Mack, J., Bell, C. D., Paterson, A.H., Zheng, C., Sankoff, D., dePamphilis, C. W., Wall, P. K., & Soltis, P.S. Polyploidy and angiosperm diversification. *American Journal of Botany* 96, 336–348 (2009).

21. McInerny, G. J. & Etienne, R. S. Stitch the niche - a practical philosophy and visual schematic for the niche concept. *Journal of Biogeography* 39, 2103–2111 (2012).

22. Leibold, M. A., Ecology, S. & Jul, N. The Niche Concept Revisited : Mechanistic Models and Community Context. *Ecology* 76, 1371–1382 (2016).

23. Colwell, R. K. & Rangel, T. F. Hutchinson's duality: The once and future niche. *Proceedings of the National Academy of Sciences of the United States of America* 106, 19651–19658 (2009).

24. Warren, D. L. In defense of 'niche modeling'. *Trends in Ecology and Evolution* 27, 497–500 (2012).

25. Allouche, O., Tsoar, A., & Kadmon, R. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* 43, 1223–1232 (2006).

26. Warren, D. L., Glor, R. E. & Turelli, M. Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *The Society for the Study of Evolution* 62, 2868-2883 (2008).

27. Peterson, A. Uses and requirements of ecological niche models and related distributional models. *Biodiversity Informatics* 3, 59–72 (2006).

28. Ownbey, M., Aase, H. Cytotaxonomic Studies in Allium I: The *Allium canadense* Alliance. Research Studies of the State College of Washington. 1-106 (1955).

29. Muscarella, R., Galante, P.J., Soley-Guardia, M., Boria, R.A., Kass, J., Uriarte, M. & Anderson, R.P. (2014). ENMeval: A R package for conducting spatially independent evaluations and estimating optimal model complexity for ecological niche models. *Methods in Ecology and Evolution*

30. Husband, B.C., Baldwin, S.J., & Suda, J. The Incidence of Polyploidy in Natural Plant Populations: Major Patterns and Evolutionary Processes. *The Plant Genome Diversity* 2(16): 255-276 (2013).

31. Laport, R.G., Hatem, L., Minckley, R.L., & Ramsey, J. Ecological niche modeling implicates climatic adaptation, competitive exclusion, and niche conservatism among *Larrea tridentata* cytotypes in North American deserts. *The Journal of the Torrey Botanical Society* 140(3):349-363 (2013).

32. Hijmans, R. J., & J. Elith.(2017, January 8). *Species Distribution Modeling in R*. Retrieved from https://cran.r-project.org/web/packages/dismo/vignettes/sdm.pdf

33. Geange, S. W., Pledger, S., Burns, K. C., & J. S. Shima. A unified analysis of niche overlap incorporating data of different types. *Methods in Ecology and Evolution* 2:175-184 (2011).

34. Glor, R.E., & D.L. Warren. (2011). Testing the ecological basis of biogeographic boundaries. *Evolution* 65:673-683.

35. Nakazato, T., D.L. Warren, & L.C. Moyle. (2010). Ecological and geographic modes of species divergence in wild tomatoes. *American Journal of Botany* 97:680-693.

36. Warren, D.L., R. E. Glor, & M. Turelli. (2008). Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62:2868-2883.

37. Warren, D.L., R.E. Glor, & M. Turelli. (2010). ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* 33:607-611.

38. Warren, D.L., & S.N. Seifert. 2011. Environmental niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications* (doi: 10.1890/10-1171.1)

39. Phillips, S. J., R. P. Anderson, & R. E. Schapire. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190:231-259.

40. Hernandez, P. A., Graham, C. H., Master, L. L., & D. L. Albert. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29: 773-785.

41. Merow, C., Smith, M. J., & J. A. Silander Jr. (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* 36: 1058–1069.

42. Pulliam, H.R. On the relationship between niche and distribution. (2000). *Ecology Letters* 3(4): 349–361.

43. Holt, R. D. (2009). Bringing the Hutchinsonian Niche into the 21st Century: Ecological and Evolutionary Perspectives. *National Academy of Science* 106: 19659-19665.

44. Ownbey, M., & H. C. Aase. (1955). Cytotaxonomic studies in Allium L.: the Allium canadense alliance, Research Studies of the State College of Washington, 1-106. State College of Washington.

45. Yong, C., DeWalt, R. E., Robinson, J. L., Tweddale, T., Hinz, L., & M. Pessino. (2013). Using Maxent to model the historic distributions of stonefly species in Illinois streams: The effects of regularization and threshold selections. *Ecological Modelling* 259 (2013) 30– 39.

46. Merow, C., Smith, M. J., & J. A. Silander, Jr. (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* 36: 1058–1069.

47. S. Phillips, Species' Distribution Modeling for Conservation Educators and Practitioners. (2010). Exercise. American Museum of Natural History, Lessons in Conservation. Available at http://ncep.amnh.org/linc

48. Baack, E.J. & Stanton, M.L. (2005) Ecological factors influencing tetraploid speciation in snow buttercups (*Ranunculus adoneus*): niche differentiation and tetraploid establishment. *Evolution* 59, 1936–1944.

49. Levin, D.A. (2003) The ecological transition in speciation. *New Phytologist* 161, 91–96.

50. PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu, created 25 Aug 2016.

51. Elith, J. (2002). Quantitative methods for modeling species habitat: Comparative performance and an application to Australian plants. In S. Ferson, & M. Burgman (Eds.), Quantitative methods for conservation biology (pp. 39–58). New York, NY: Springer-Verlag.

52. Phillips, S. J., & Dudik, M. (2008). Modeling of species distributions with e: New extensions and a comprehensive evaluation. *Ecography*, 31, 161–175.

53. Walters AD, Ford D, Chong ET, et al. High-resolution ecological niche modelling of threatened freshwater mussels in east Texas, USA. Aquatic Conserv: Mar Freshw Ecosyst. (2017);0:1–11. https://doi.org/10.1002/aqc.2784

54. S.D Prince, S.J Goetz, R.O Dubayah, K.P Czajkowski, M Thawley, Inference of surface and air temperature, atmospheric precipitable water and vapor pressure deficit using Advanced Very High-Resolution Radiometer satellite observations: comparison with field observations, In Journal of Hydrology, Volumes 212–213, (1998), Pages 230-249, ISSN 0022-1694, https://doi.org/10.1016/S0022-1694(98)00210-8.

55. Hubbard, K. G., R. Mahmood, & C. Carlson. 2003. Estimating Daily Dew Point Temperature for the Northern Great Plains Using Maximum and Minimum Temperature. Agron. J. 95:323-328. doi:10.2134/agronj2003.3230

56. Banta, J. A., I. M. Ehrenreich, S. Gerard, L. Chou, A. Wilczek, J. Schmitt, P. X. Kover, & M. D. Purugganan. (2012). Climate envelope modeling reveals intraspecific relationships among flowering phenology, niche breadth, and potential range size in *Arabidopsis thaliana*. *Ecology Letters* 15: 769-777